

Open Research Online

The Open University's repository of research publications and other research outputs

Genetic diversity and evolution of hepatitis C virus

Thesis

How to cite:

Harris, Kathryn Ann (2000). Genetic diversity and evolution of hepatitis C virus. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2000 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000e2c5>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Genetic Diversity and Evolution of Hepatitis C Virus

Submitted in fulfilment of the conditions
governing candidates for the degree of

Doctor of Philosophy

of

The Open University

Discipline: Life Sciences

Kathryn Ann Harris

Hepatitis and Retrovirus Laboratory

Central Public Health Laboratory

61 Colindale Avenue

London NW9

November 1999

AUTHOR'S No: R468379X

DATE OF AWARD: 31 MARCH 2000

Abstract

Inter- and intra-host HCV variation was investigated. First, a polymerase chain reaction-restriction fragment length polymorphism procedure was used to assign genotypes and subtypes to HCV infecting 567 individuals (comprising haemophilia patients, blood donors, intravenous drug users, attenders of antenatal and genito-urinary medicine clinics and chronic liver disease patients) from England and Wales. The majority of infections were associated with types 3a, 1a and 1b, and genotype distributions were generally similar in different sub-populations. Only 1% of individuals were identified as being infected with more than one subtype. The intra-host variability of HCV in a selection of haemophilia patients, blood donors and intravenous drug users was then studied. For each individual, PCR clones derived from the NS5b and 5' non-coding regions of the HCV genome were screened for sequence differences by denaturing gradient gel electrophoresis (DGGE) and nucleotide sequencing. The complexity and diversity of HCV quasispecies, though differing between individuals, could not be correlated with the risk group to which the study patients belonged. Furthermore, no mixed genotype or subtype infections were identified. Thus the hypothesis that multiply exposed individuals are infected with a greater variety of HCV variants could not be substantiated. The DGGE procedure was further used to investigate the hypothesis that HCV genetic evolution occurs uniformly in patients during the acute phase of infection. Changes in diversity in the HCV hypervariable region 1 in individuals undergoing seroconversion were observed to differ between patients, thereby negating that hypothesis. Moreover, in a given individual, HCV could be subjected to either positive or negative selective pressure. Thus, factors other than the acute-phase host response determine the course of HCV genetic evolution.

Acknowledgements

Thank you to everybody that has helped me produce this thesis, in particular my supervisors, Dr. Chong-Gee Teo and Dr. Vincent Emery.

Thanks also to my colleagues at CPHL for their support and friendship over the last three years.

Abbreviations

ALT	Alanine aminotransferase
ANC	Antenatal clinic
bDNA	Branched DNA
bp	Base pair
cDNA	Complementary DNA
CTL	Cytotoxic T-lymphocyte
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
dn	Non-synonymous change
ds	Synonymous change
E	Envelope
EDTA	Disodium ethylenediaminetetra-acetic acid
EIA	Enzyme immunosorbent assay
ELISA	Enzyme linked immunosorbent assay
GD	Genetic distance
GUM	Genito-urinary medicine
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HMA	Heteroduplex mobility assay
HVR	Hypervariable region
IDU	Intravenous drug user
IRES	Internal ribosomal entry site
ISDR	Interferon sensitivity determining region
MMLV	Murine Maloney leukaemia virus
mRNA	Messenger RNA
NANB	Non-A non-B
NCR	Non-coding region
NS	Non-structural
ORF	Open reading frame

PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PHLS	Public Health Laboratory Service
PKR	Protein kinase
PTB	Polypyrimidine tract binding protein
RdRp	RNA-dependent RNA polymerase
RFLP	Restriction fragment length polymorphism
RIBA	Recombinant Immunoblot Assay
RNA	Ribonucleic acid
RT	Reverse transcription
Sn	Shannon entropy (normalised)
SSCP	Single-strand conformational polymorphism
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
Ts	Transition
Tv	Transversion
UTR	Untranslated region
UV	Ultraviolet

Table of Contents

Abbreviations	4
List of figures	12
List of tables	17
Chapter 1 Introduction	19
1.1 Discovery	20
1.2 Disease association	21
1.3 Epidemiology	22
1.4 Genetic organisation and expression	22
1.4.1 The non-coding regions	23
1.4.2 The core protein	26
1.4.3 The envelope proteins	27
1.4.4 The non-structural proteins	27
1.5 Genetic variability	28
1.6 Genotypes	29
1.6.1 Assigning HCV genotyping	29
1.6.2 Geographical distribution of HCV genotypes	30
1.6.3 HCV genotypes and disease	31
1.7 The quasispecies nature of HCV	33
1.7.1 Quasispecies and persistent infection	33
1.7.2 Quasispecies and progression of liver disease	34
1.7.3 Quasispecies and acute infection	35
1.7.4 Quasispecies and the response to antiviral therapy	36
1.8 HCV quasispecies and PCR studies	36

1.9 Quasispecies assessment by single-strand conformational polymorphism analysis..	37
1.10 Quasispecies analysis using the heteroduplex mobility assay	38
1.11 Quasispecies analysis using denaturing gradient gel electrophoresis	39
1.12 Aims of this study	42
Chapter 2 Materials and Methods.....	43
2.1 Amplification of HCV 5' Non-Coding Region (NCR).....	44
2.1.1 RNA Extraction.....	44
2.1.2 Synthesis of complementary DNA (cDNA)	44
2.1.3 Nested Polymerase chain reaction (PCR)	44
2.1.4 Agarose gel electrophoresis	45
2.2 Genotyping of HCV using the 5'NCR.....	45
2.2.1 Restriction digests of second round PCR product	45
2.2.2 Polyacrylamide gel electrophoresis (PAGE)	46
2.2.3 Sequencing	46
2.2.4 Cloning.....	47
2.3 Statistical analysis	47
2.4 Amplification of HCV Non-Structural Region 5b (NS5b)	48
2.4.1 RNA extraction and cDNA synthesis	48
2.4.2 Single round PCR.....	48
2.5 Genotyping of HCV using the NS5b region	49
2.5.1 Gel purification of PCR products.....	49
2.5.2 Cloning and sequencing of purified PCR products.....	49
2.6 Analysis of HCV quasispecies by denaturing gradient gel electrophoresis (DGGE)	49
2.6.1 Single round HCV 5'NCR PCR.....	50
2.6.2 Single round HCV NS5b PCR.....	50

2.6.3 Gel purification of HCV NS5b PCR products	50
2.6.4 Amplification of HCV envelope region, incorporating HVR-1.....	50
2.6.5 Cloning.....	51
2.6.6 Colony PCR to amplify 5'NCR inserts and the attachment of a GC-clamp.....	52
2.6.7 Colony PCR to amplify NS5b inserts and the attachment of a GC-clamp	52
2.6.8 Colony PCR to amplify envelope/HVR1 inserts and the attachment of a GC-clamp	52
2.6.9 Preparation of a denaturing gradient polyacrylamide gel for simultaneous analysis of 5'NCR and NS5b colony PCR products	53
2.6.10 Preparation of a denaturing gradient polyacrylamide gel for analysis of envelope/HVR-1 colony PCR products.....	53
2.6.11 Loading and electrophoresis	54
2.6.12 Gel staining	54
2.6.13 Sequencing of variants	54
2.6.14 Analysis of sequence data	55
2.6.15 Calculating genetic complexity and diversity	55
2.6.16 Statistical analysis	56
Chapter 3 Distribution of HCV Genotypes in England and Wales.....	57
3.1 Introduction	58
3.2 Materials and Methods.....	59
3.2.1 Patients	59
3.2.2 Methods.....	59
3.3 Results.....	62
3.3.1 Genotype distribution.....	62
3.3.2 Statistical analysis	66

3.4 Discussion	66
3.4.1 Genotyping techniques.....	66
3.4.2 Prevalence of 3a and mixed infections in other studies	68
3.4.3 Genotype and mode of infection.....	70
3.4.4 Implications of HCV genotype distribution in England and Wales.....	71
Chapter 4 Investigating HCV Quasispecies Using Denaturing Gradient Gel Electrophoresis	73
4.1 Introduction.....	74
4.2 Materials and Methods.....	76
4.2.1 Patients	76
4.2.2 Method	76
4.2.3 Development of methodology.....	77
4.2.3.1 Selection of suitable gradient for analysis of the 5'NCR.....	77
4.2.3.2 DGGE analysis of the hyper-variable region 1 (HVR1).....	79
4.2.3.3 DGGE analysis of the NS5b region	80
4.2.3.4 Electrophoresis conditions	81
4.2.3.5 Simultaneous analysis of both genomic regions	81
4.2.3.6 Investigation of heteroduplex bands	82
4.3 Results.....	85
4.3.1 DGGE and nucleotide sequences	85
4.3.2 Summary of diversity and sequence data.....	106
4.3.3 Genotyping clonotypes by phylogenetic analysis	111
4.4 Discussion	125
4.4.1 Quasispecies complexity and diversity	125
4.4.2 Mode of acquisition and diversity	127

4.4.3 Ratio of synonymous to non-synonymous changes	127
4.4.4 Genotype and subtypes	128
4.4.5 Infection with multiple genotypes or subtypes	129
4.4.6 Sources of artefactual error	132
Chapter 5 HCV Quasispecies Evolution During Seroconversion.....	135
5.1 Introduction	136
5.2 Materials and Methods	140
5.2.1 Patients	140
5.2.2 Methods.....	142
5.3 Results.....	143
5.3.1 HVR-1 quasispecies during acute infection.....	143
5.3.2 Patient RZ	143
5.3.3 Patient 11.....	147
5.4 Discussion	159
5.4.1 Variation in HCV diversity in the early phase of acute infection.....	159
5.4.2 Variation in HCV evolution during the course of acute infection	160
5.4.3 Limitations of experimental methods	163
Chapter 6 Summary and Future Studies.....	165
6.1 Summary of thesis.....	166
6.1.1 Overview of experimental design	166
6.1.2 Conclusions.....	167
6.1.2.1 Distribution of HCV genotypes in England and Wales	167
6.1.2.2 Infection with multiple HCV genotypes or subtypes is rare	168
6.1.2.3 The direction of evolution of HCV during acute infection is variable	168
6.2 Future studies	169

References..... 170

Publications and presentations relating to this thesis..... 194

List of figures

- Figure 1.1: Organisation of the HCV genome.
- Figure 1.2: Proposed secondary structure of the 5'NCR.
- Figure 1.3: Schematic diagram of mutation detection by SSCP analysis.
- Figure 1.4: Schematic diagram of mutation detection by DGGE analysis.
- Figure 3.1: Restriction fragment length polymorphisms in the 174-bp fragment of the 5'NCR.
- Figure 3.2: A developmental genotyping gel.
- Figure 3.3: A typical genotyping gel.
- Figure 3.4: Overall genotype distribution in 567 individuals.
- Figure 3.5: Genotype distribution among 77 haemophilia patients.
- Figure 3.6: Genotype distribution among 68 blood donors.
- Figure 3.7: Genotype distribution among 78 intravenous drug users.
- Figure 3.8: Genotype distribution among 62 ANC attenders.
- Figure 3.9: Genotype distribution among 86 GUM clinic attenders.
- Figure 3.10: Genotype distribution among 194 liver disease patients.
- Figure 4.1: Developmental DGGE analysis of the 5'NCR using a 0-80% gradient.
- Figure 4.2: Developmental DGGE analysis of the 5'NCR using a 30-70% gradient.
- Figure 4.3: Developmental DGGE analysis of the HVR-1.
- Figure 4.4: Developmental DGGE analysis of the NS5b using 35 colony PCR cycles.
- Figure 4.5: Developmental DGGE analysis of the NS5b using 20 colony PCR cycles.
- Figure 4.6: Simultaneous DGGE analysis of the 5'NCR and the NS5b regions.
- Figure 4.7: Detection of multiple bands on DGGE analysis of a colony PCR product.

Figure 4.8: Investigation of multiple bands in DGGE analysis.

Figure 4.9: DGGE analysis of clonotypes infecting patient H069.

Figure 4.10: DGGE analysis of clonotypes infecting patient H071.

Figure 4.11: DGGE analysis of clonotypes infecting patient H075.

Figure 4.12: DGGE analysis of clonotypes infecting patient H858.

Figure 4.13: DGGE analysis of clonotypes infecting patient H865.

Figure 4.14: DGGE analysis of clonotypes infecting patient BD244.

Figure 4.15: DGGE analysis of clonotypes infecting patient BD259.

Figure 4.16: DGGE analysis of clonotypes infecting patient BD268.

Figure 4.17: DGGE analysis of clonotypes infecting patient BD424.

Figure 4.18: DGGE analysis of clonotypes infecting patient BD426.

Figure 4.19: DGGE analysis of clonotypes infecting patient IDU189.

Figure 4.20: DGGE analysis of clonotypes infecting patient IDU191.

Figure 4.21: DGGE analysis of clonotypes infecting patient IDU230.

Figure 4.22: DGGE analysis of clonotypes infecting patient IDU240.

Figure 4.23: DGGE analysis of clonotypes infecting patient IDU323.

Figure 4.24: Amino acid sequences of the NS5b clonotypes infecting patient H069.

Figure 4.25: Amino acid sequences of the NS5b clonotypes infecting patient H071.

Figure 4.26: Amino acid sequences of the NS5b clonotypes infecting patient H075.

Figure 4.27: Amino acid sequences of the NS5b clonotypes infecting patient H858.

Figure 4.28: Amino acid sequences of the NS5b clonotypes infecting patient H865.

Figure 4.29: Amino acid sequences of the NS5b clonotypes infecting patient BD244.

Figure 4.30: Amino acid sequences of the NS5b clonotypes infecting patient BD259.

Figure 4.31: Amino acid sequences of the NS5b clonotypes infecting patient BD268.

- Figure 4.32: Amino acid sequences of the NS5b clonotypes infecting patient BD424.
- Figure 4.33: Amino acid sequences of the NS5b clonotypes infecting patient BD426.
- Figure 4.34: Amino acid sequences of the NS5b clonotypes infecting patient IDU189.
- Figure 4.35: Amino acid sequences of the NS5b clonotypes infecting patient IDU191.
- Figure 4.36: Amino acid sequences of the NS5b clonotypes infecting patient IDU230.
- Figure 4.37: Amino acid sequences of the NS5b clonotypes infecting patient IDU240.
- Figure 4.38: Amino acid sequences of the NS5b clonotypes infecting patient IDU323.
- Figure 4.39: Analysis of haemophilia patient 5'NCR sequences with CLUSTAL V.
- Figure 4.40: Analysis of haemophilia patient 5'NCR sequences with FITCH.
- Figure 4.41: Analysis of blood donor 5'NCR sequences with CLUSTAL V.
- Figure 4.42: Analysis of blood donor 5'NCR sequences with FITCH.
- Figure 4.43: Analysis of IDU 5'NCR sequences with CLUSTAL V.
- Figure 4.44: Analysis of IDU 5'NCR sequences with FITCH.
- Figure 4.45: Analysis of haemophilia patient NS5b sequences with CLUSTAL V.
- Figure 4.46: Analysis of haemophilia patient NS5b sequences with FITCH.
- Figure 4.47: Analysis of blood donor NS5b sequences with CLUSTAL V.
- Figure 4.48: Analysis of blood donor NS5b sequences with FITCH.
- Figure 4.49: Analysis of IDU NS5b sequences with CLUSTAL V.
- Figure 4.50: Analysis of IDU NS5b sequences with FITCH.
- Figure 5.1: HVR-1 quasispecies distribution in a chronically infected individual with liver dysfunction and a post-liver transplant patient.
- Figure 5.2: Diagrammatic representation of a genetic bottleneck.
- Figure 5.3: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient RZ at time point 1.

- Figure 5.4: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient RZ at time point 2.
- Figure 5.5: Nucleotide sequences of envelope region/HVR-1 clonotypes infecting Patient RZ.
- Figure 5.6: Amino acid sequences of envelope region/HVR-1 clonotypes infecting Patient RZ.
- Figure 5.7: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 1.
- Figure 5.8: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 2.
- Figure 5.9: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 3.
- Figure 5.10: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 4.
- Figure 5.11: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 5.
- Figure 5.12: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 11 at time point 6.
- Figure 5.13: Nucleotide sequences of envelope region/HVR-1 clonotypes infecting Patient 11.
- Figure 5.14: Amino acid sequences of envelope region/HVR-1 clonotypes infecting Patient 11.
- Figure 5.15: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 1.

- Figure 5.16: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 2.
- Figure 5.17: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 3.
- Figure 5.18: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 4.
- Figure 5.19: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 5.
- Figure 5.20: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient 14 at time point 6.
- Figure 5.21: Nucleotide sequences of envelope region/HVR-1 clonotypes infecting Patient 14.
- Figure 5.22: Amino acid sequences of envelope region/HVR-1 clonotypes infecting Patient 14.
- Figure 5.23: Dendrogram to show relatedness of HVR-1 variants infecting Patient 14.
- Figure 5.24: Chart to show how viral load, ALT values, antibody levels and ds/dn ratios change during acute infection in Patient 11.
- Figure 5.25: Chart to show how viral load, ALT values, antibody levels and ds/dn ratios change during acute infection in Patient 14.

List of tables

- Table 3.1: Comparison of HCV genotype distributions in eight countries.
- Table 3.2: Prevalence of mixed genotype or subtype infections using different genotyping methods.
- Table 4.1: Sn, mean GD and Ts/Tv for 5'NCR quasispecies infecting 15 individuals.
- Table 4.2: Sn, mean GD, Ts/Tv and ds/dn for NS5b quasispecies infecting 15 individuals.
- Table 4.3: Sn and mean GD for translated NS5b quasispecies infecting 15 individuals.
- Table 4.4: Ranges and mean values for Sn and mean GD for 5'NCR quasispecies infecting 15 individuals analysed collectively and according to patient group.
- Table 4.5: Ranges and mean values for Sn, mean GD and ds/dn for NS5b quasispecies infecting 15 individuals as a whole and as three patient groups.
- Table 4.6: Ranges and mean values for Sn and mean GD for translated NS5b quasispecies infecting 15 individuals as a whole and as three patient groups.
- Table 4.7: Comparison of genotypes assigned by RFLP analysis and sequence analysis of 5'NCR and NS5b clones.
- Table 5.1: Clinical information for serial samples taken from Patients RZ.
- Table 5.2: Clinical information for serial samples taken from Patients 11.
- Table 5.3: Clinical information for serial samples taken from Patients 14.
- Table 5.4: Proportion of synonymous (ds) and non-synonymous (dn) changes among variants infecting Patient RZ at time points 1 and 2.

Table 5.5: Proportion of synonymous (ds) and non-synonymous (dn) changes among variants infecting Patient 11 at time points 1 to 6.

Table 5.6: Proportion of synonymous (ds) and non-synonymous (dn) changes among variants infecting Patient 14 at time points 1 to 6.

Chapter 1

Introduction

1.1 Discovery

When in the late 1970s blood donations began to be screened for antibodies to and markers of hepatitis viruses A and B, it was observed that approximately 10% of blood transfusion recipients continued to develop post-transfusion hepatitis. When all other factors known to cause hepatitis had been excluded, the term non-A non-B (NANB) hepatitis was used to refer to this condition. The agent responsible remained elusive for almost two decades. It is now known that hepatitis C virus (HCV) (Choo et al. 1989) is the major causative agent of NANB hepatitis, accounting for about 80% of such cases.

The discovery of HCV (Choo et al. 1989) was an important scientific breakthrough. The virus was identified solely by the isolation and characterisation of its genome. Plasma pools from chimpanzees infected with serum taken from patients with NANB hepatitis were used as starting material from which total nucleic acid was extracted and reverse transcribed. The resulting complementary DNA (cDNA) was used to construct an expression library, which was screened with sera derived from patients with NANB hepatitis. More than one million clones were screened before one was found to react with sera from several patients and from the infected chimpanzees. Three further over-lapping clones were isolated and a 1089-nucleotide continuous open reading frame (ORF) was reconstructed (Choo et al. 1989; Booth, 1998a). This ORF formed the basis of the antigen (c100-3) used in the first generation enzyme-linked immunosorbent assay (ELISA) (Kuo et al. 1989). The sequence of the entire genome structure of the virus soon became fully characterised (Choo et al. 1991).

More recently HCV has been visualised by electron microscopy (Kaito et al. 1994) and second and third generation serological tests have been developed (Courouce et al. 1994; Barrera et al. 1995). However, to this day the propagation of HCV *in vitro* has still not been fully achieved, since no cell lines have been found to be fully permissive to viral replication. The replication of the virus therefore remains poorly understood (Kato et al. 1998; Fournier et al. 1998; Fipaldini et al. 1999). It is likely to resemble that used by the closely related flaviviruses, which involves the synthesis of a negative strand followed by transcription of the genome from a double stranded intermediate (Booth, 1998a).

1.2 Disease association

HCV can cause acute hepatitis, but it is chronic infection that presents the greatest public health challenge. There is no vaccine against HCV and current treatment regimes, by use of interferon alpha and ribavirin, are effective in only about a third of cases (Dusheiko et al. 1996; Davis et al. 1998; Cohen, 1999). Following acute infection, the majority (80%) of individuals become chronically infected and 20% of these individuals will eventually develop liver cirrhosis, and in some cases, hepatocellular carcinoma (Farci et al. 1997; Idilman et al. 1998; Booth, 1998a). It is likely that the immune response, in particular the cytotoxic T-lymphocyte (CTL) response, is responsible for much of the liver damage seen in chronic infection, the infected hepatocytes being destroyed by components of the cellular immune system such as cytokines and macrophages (Koziel, 1997).

Chronic HCV infection also seems to be associated with a number of extrahepatic manifestations, in particular some autoimmune diseases such as type II cryoglobulinaemia (Dumoulin et al. 1997; Lunel and Musset, 1998) and non-Hodgkin's lymphoma (Dammacco et al. 1998).

1.3 Epidemiology

The World Health Organisation estimates that 170 million people worldwide are infected with HCV (Booth, 1998a). An estimated 200,000 – 400,000 of these infections are in the U.K. The prevalence of the virus in healthy blood donors ranges from 0.01 - 0.02% in the UK and northern Europe to 10% in parts of Africa (Booth, 1998a).

In the majority of cases, HCV transmission is via parenteral routes such as intravenous drug use, or transfusion with blood or blood products. Since the introduction of screening of blood donations, intravenous drug use is now the most significant risk factor for acquisition of HCV infection. In the U.K., the prevalence of HCV among intravenous drug users (IDUs) may be as high as 70% (Ramsay et al. 1998). Other parenteral routes relate to haemodialysis, organ transplant, piercing and tattooing. For a large number of individuals with chronic HCV infection no obvious route of transmission can be found. This may be due to less obvious routes of parenteral transmission, such as use of shared toothbrushes or razors (Brackmann et al. 1993; Booth, 1998a). Non-parenteral transmission may also occur. There is evidence to suggest that in some cases HCV may be sexually or vertically transmitted (Pipan et al. 1996). Such events are relatively rare and may be dependent on additional factors, such as co-infection with HIV (Wright et al. 1994; Booth, 1998a).

1.4 Genetic organisation and expression

Sequence analysis of the genome has led to the classification of HCV as a member of the *Flaviviridae* family. Overall, the sequence homology between HCV and other flaviviruses is relatively low but they share a similar genomic organisation. The HCV genome is a single-stranded, positive sense RNA molecule of approximately 9,500 nucleotides in length. It codes for a single polyprotein of about 3,000 amino acids that is post

translationally cleaved into structural and non-structural proteins. The nomenclature of these proteins and the order in which the genes lie in the single open reading frame (ORF) are shown in Figure 1.1.

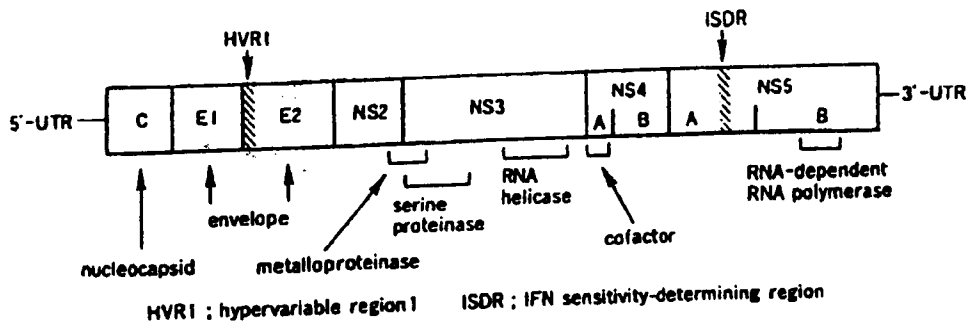


Figure 1.1: Organisation of the HCV genome. There are two non-coding regions (NCR) that flank the open reading frame. The polyprotein comprises three structural genes that code for the core protein (C) and the two envelope glycoproteins (E1 and E2), plus four nonstructural (NS) proteins. Diagram taken from Hayashi et al. (1999).

1.4.1 The non-coding regions

Two highly conserved non-coding regions (NCR), which are also referred to in the literature as untranslated regions (UTR), flank the ORF. The ORF codes for the core protein (C), two envelope proteins (E1 and E2) and four non-structural proteins (NS2 – 5).

The 5'NCR consists of 341 nucleotides that are predicted to form secondary stem-loop structures. The 5'NCR may function as an internal ribosomal entry site (IRES) (Smith et al. 1995; Major and Feinstone, 1997; Hellen and Pestova, 1999). The IRES element directs cap-independent translation, a mechanism first described for picornaviruses. For the

majority of eukaryotes, translation is initiated by the binding of a 43S ribosomal complex to the capped 5' end of the mRNA. Upon binding, eukaryotic initiation factors (eIFs) act to unwind the RNA to form the binding site and promote the binding of the 43S complex. Following RNA unwinding, the initiation complex scans downstream to form a stable 48S complex with the first AUG triplet which then assumes its role as the initiation codon. In HCV and picornaviruses, however, the IRES element recruits the 43S complex to the precise location of the initiation codon on the mRNA to form the 48S complex. This mechanism does not require all of the eIFs or the ribosomal scanning for cap-dependent translation (Hellen and Pestova, 1999). Translation of HCV RNA is inefficient compared to other flaviviruses, such as yellow fever virus, that use the more common cap-dependent translation mechanism (Malet et al. 1998). In evolutionary terms, it is unclear why HCV requires an IRES element within its 5'NCR, but it is possible that it functions to decrease translation efficiency, facilitating viral persistence (Malet et al. 1998). Figure 1.2 shows the proposed secondary structure of the 5'NCR.

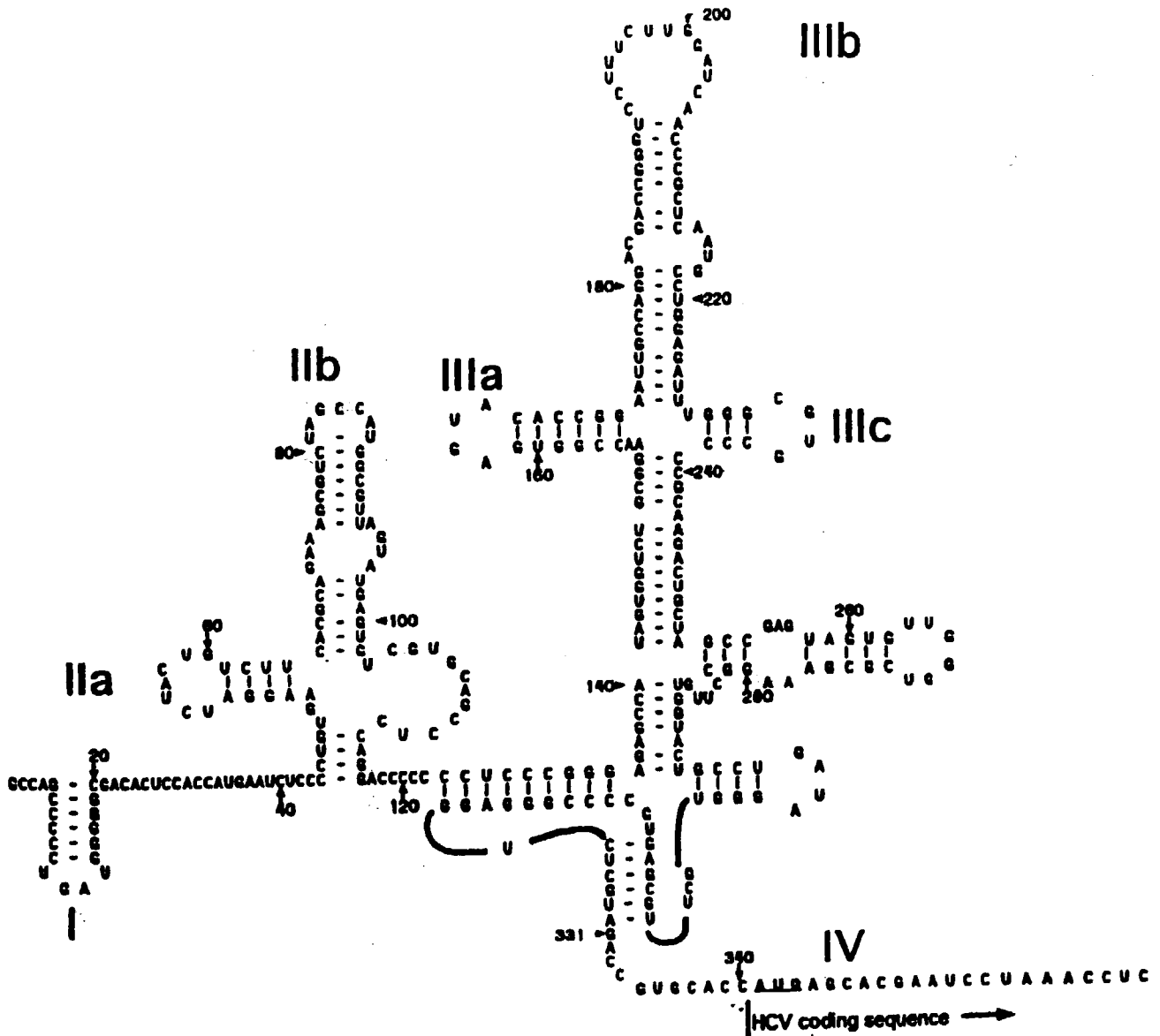


Figure 1.2: Proposed secondary structure of the 5'NCR that incorporates an IRES element. There are four domains of secondary structure, labelled I – IV. The initiation codon is underlined. Diagram taken from Hellen and Pestova (1999).

The 3'NCR may also have a role in the regulation of translation. The majority of eukaryotic mRNAs and many viral RNAs contain a stretch of A residues at their 3' end, called a poly (A) tail, which enhances translation. HCV does not have a poly (A) tail; instead its 3'NCR region contains a stretch of 98 nucleotides, known as the X-region. This region assumes a highly conserved stem-loop structure, which binds a polypyrimidine tract-binding protein

(PTB) (Ito and Lai, 1997). The hairpin structure in this region may act as the primer site necessary for negative strand synthesis during replication. The IRES element also binds PTB, and interactions between the 5'NCR and the 3'NCR, via this protein, may also regulate HCV translation (Ito et al. 1998).

1.4.2 The core protein

The core protein is 191 amino acids in length and highly conserved. It is immunogenic, containing several linear B-cell epitopes. Synthetic peptides and peptides expressed from this region have been used extensively as antigens for the detection of HCV antibodies in patient sera (Major and Feinstone, 1997). The core protein is also weakly oncogenic (Chang et al. 1998). The core gene is expressed as three different proteins of 21 kDa, 19 kDa and 16kDa molecular weight. The smallest of these shows a predominantly nuclear localisation *in vitro*. It may act to regulate host gene expression, in particular cellular proto-oncogenes and tumour suppressors. The ratio of the three core gene products *in vivo* is likely to be important for the transforming potential of the core. The distribution of core protein in the nuclei of hepatocytes in natural infection is not known. There are no established mechanisms by which the core is oncogenic, although recent studies are beginning to uncover interactions of the core protein with various cellular pathways. It has been shown to give cells a growth advantage when expressed in cell lines carrying the oncogene v-H-ras (Tsuchihara et al. 1999) and it represses the transcriptional activity of p21, a promoter that regulates the cell cycle (Ray et al. 1998b). The core protein has also been shown to inhibit apoptotic cell death (Ray et al. 1996; Ray et al. 1998a; Marusawa et al. 1999).

1.4.3 The envelope proteins

The glycoproteins E1 and E2 form the envelope of the virion. A region of the E2 protein displays an unusually high degree of amino acid sequence diversity and is referred to as hypervariable region 1 (HVR 1). There is evidence to suggest that the HVR1 is presented on the surface of the virion and is the probable site at which neutralising antibodies are directed (Farci et al. 1996; Lechner et al. 1998). The extreme degree of sequence heterogeneity in this region probably contributes to the ability of HCV to escape from the host immune response. E2 inhibits the kinase activity of the interferon-inducible protein kinase (PKR), a cellular kinase, and blocks its inhibitory effect on protein synthesis and cell growth. This interaction between E2 and PKR may be one mechanism by which HCV circumvents the antiviral effects of interferon (Taylor et al. 1999).

1.4.4 The non-structural proteins

The NS3 protein possesses serine protease activity. It cleaves the polyprotein at the NS2/NS3 border. The NS2 protein acts as a co-factor for this process. NS3 then cleaves at the following borders sequentially: NS3/NS4A, NS5A/NS5B, NS4B/NS5A and NS4A/NS4B (Major and Feinstone, 1997). The NS3 protein also functions as an RNA helicase, with the ability to unwind RNA:DNA and RNA:RNA duplexes (Gwack et al. 1997). As helicase activity is essential for viral replication, the NS3 protein is an attractive target for antiviral drugs (Bartenschlager, 1997; Littlejohn et al. 1998).

The NS4 region codes for two proteins: NS4A, which acts as a cofactor for NS3 protease activity; and NS4B, whose function is unknown.

The NS5 region also codes for two proteins. The primary function of NS5A is not clear but it appears to be a potent transcriptional activator (Kato et al. 1997; Fukuma et al. 1998). Certain mutations in NS5A influence the sensitivity of HCV to treatment with interferon. This part of the NS5A protein, between amino acids 237 and 276 of the 447 amino acid protein, is known as the 'interferon sensitivity determining region' (ISDR). The mechanism by which the ISDR affects response to IFN is still unknown. It could be via interaction with PKR. Mutations in the ISDR may affect the binding of NS5A to PKR (Polyak et al. 1999).

The NS5B protein is an RNA dependent RNA polymerase (RdRp). It shows homology with other RdRps, and an expressed NS5B protein has been shown to possess RdRp activity *in vitro* (Behrens et al. 1996; Lohmann et al. 1997). The predicted structure of the HCV RdRp is similar to other RNA polymerases (Lohmann et al. 1998). Several well-conserved structural motifs may be essential for enzymatic activity. The key role of NS5b in viral replication makes it another candidate for antiviral therapy (Bartenschlager, 1997; Littlejohn et al. 1998).

1.5 Genetic variability

The HCV genome displays a high level of genetic variability. This is a result of error prone replication by the viral polymerase. Most viruses possessing an RNA polymerase lack a proof reading function. In the absence of methods for virus classification that depend on *in vitro* culture systems, classification systems based entirely on nucleotide sequence have been proposed for HCV. The most widely accepted of these is that proposed by Simmonds et al. (1993). HCV can be divided into six major genotypes (1 - 6) and each of these comprises a number of subtypes, denoted by lower case letters. Thus genotypes/subtypes are designated '1a', '2b', '3a' etc. HCV genomes belonging to different genotypes share

66% - 69% sequence similarity and those belonging to the same genotype but having different subtypes share 77% - 80% sequence similarity. Most isolates belonging to the same subtype share 91% - 98% similarity.

Knowledge of the rate at which the HCV genome accumulates mutations enables estimates to be made about when the genotypes, and later the subtypes, diverged from each other. The 'molecular clock' for HCV has been estimated to progress at a rate of 0.144% of the complete genome per year. Using this figure the major genotypes of HCV were initially thought to have diverged around 100 years ago, making it a relatively recent viral infection of humans. However, this figure may be an underestimate since the molecular clock assumes a constant mutation rate. In reality the mutations observed during longitudinal infection studies are heavily biased toward silent mutations, those causing a change in the amino acid sequences being more likely to be detrimental to the virus and therefore selected against. Taking this into account, and the possibility of 'multiple hits' (when mutations occur more than once at the same site resulting in no net increase in sequence divergence), a more realistic estimate for the divergence of the major HCV genotypes is 1,000 or even 10,000 years ago (Simmonds, 1995).

1.6 Genotypes

1.6.1 Assigning HCV genotyping

HCV isolates are classified using phylogenetic analysis of nucleotide sequences in the course of which nucleotide sequences are compared and percentage similarities calculated. Dendrograms have been produced to describe the way in which sequences are related to each other, as sequences belonging to the same genotype or subtype will cluster together.

This type of analysis requires powerful computational methods that use various mathematical algorithms to provide the most accurate picture of how nucleotide sequences relate to each other (Clewley, 1998a; 1998c). Whole genome sequences provide the most accurate basis for assigning genotypes. However, obtaining and analysing sequences of this length is too labour intensive to provide a practical solution for the classification of many HCV isolates. Therefore portions of the genome have been used instead, a 222-base pair fragment of the NS5b gene being particularly selected (Simmonds et al. 1993; Dusheiko et al. 1994a; Mellor et al. 1995). Phylogenetic analysis of gene fragments does provide accurate information about genotypes.

Several genotyping methods are based on more highly conserved regions of the genome, in particular the 5'NCR. The sequence variation between genotypes being much less in this region, many subtypes share identical sequences. However, the six major genotypes and several subtypes possess well conserved sequence polymorphisms in the 5'NCR, so methods such as restriction fragment length polymorphism (RFLP) (Davidson et al. 1995; Pohjanpelto et al. 1996), probe hybridisation and line probe assay (Lipa) (Stuyver et al. 1996) can be applied to assign genotypes. The advantages of using such methods are that they are rapid and relatively inexpensive, providing a useful tool for genotyping large numbers of HCV isolates. They are not as accurate as genotyping by sequence analysis of coding regions, many subtypes being assigned on the basis of single base differences.

1.6.2 Geographical distribution of HCV genotypes

Information about the prevalence of HCV genotypes around the world is important for understanding the origins and transmission of the virus. However, in large parts of the world (particularly the third world) little is known about which HCV genotypes prevail.

The most studied regions are the USA and Europe, where types 1a, 1b, 2a, 2b, 2c and 3a (in differing proportions) account for the majority of infections (Davidson et al. 1995; Simmonds, 1995). Crossing from Southeast Europe into Middle East countries, a dramatic change in genotype distribution is seen, genotype 4 being very common in these countries.

Both in Central Africa, where genotype 4 predominates, and West Africa, where most infections are by genotypes 1 and 2, an impressive array of subtypes are seen. Many isolates cannot be classified as existing subtypes (Ruggieri et al. 1996; Rapicetta et al. 1998; Wansbrough-Jones et al. 1998), and one study, in West Africa, identified 16 new subtypes of genotypes 1 and 2 (Jeannel et al. 1998). Such heterogeneity of subtypes suggests that HCV has been co-evolving in isolation within their hosts in these areas for a long time.

In Far Eastern countries such as Japan, Taiwan and China, types 1b, 2a and 2b are prevalent, whereas in Singapore and Thailand type 3 is common. Type 5a appears to be restricted to South Africa, and 6a is particular to Hong-Kong, Vietnam and Macau. The existence of genotypes 7, 8 and 9 has also been reported in Indonesia and Vietnam (Tokita et al. 1994; Tokita et al. 1995), but these are thought to be additional subtypes of the existing six genotypes (Simmonds et al. 1996).

1.6.3 HCV genotypes and disease

Although for some RNA viruses (e.g. polio virus or dengue virus) different types do not differ in their pathogenic characteristics, for others (such as Coxsackie B or pestiviruses) different subtypes differ considerably in the severity of disease they cause or in the organs they target (Melnick, 1996; Rice, 1996). HCV seems to belong to the latter category of

viruses, as there is some suggestion for HCV that the progress of liver disease and the response to certain antiviral drugs may be correlated with genotype.

Many studies have implicated type 1b in the development of more severe liver disease and resistance to treatment with alpha-interferon (Dusheiko and Simmonds, 1994b; Isaacson et al. 1997). However, it is difficult to be sure whether it is genotype 1b itself that determines these outcomes as this genotype is more prevalent in older age groups and would therefore be associated with longer duration of disease.

Most studies examining the mechanisms by which genotype 1b is associated with poor response to the drug have concentrated on the NS5a gene, which contains the ISDR (Enomoto et al. 1995). The ISDR appears to be located in a relatively stable region of the genome, and the variation found in mutant ISDR sequences may be geographically restricted (Rispetter et al. 1998). As mentioned earlier, mutations in this gene are thought to confer sensitivity to interferon (Fukuma et al. 1998). The 'wild-type' ISDR sequence is found in the Japanese prototype 1b strain, 'HCV-J'. Strains with one to three amino-acid mutations are 'intermediate' types and those with more than three amino acid mutations are 'mutant' types. Patients infected with wild-type strains appear not to respond well to treatment with interferon; those infected with mutant strains seem to be interferon responders (Enomoto et al. 1995; Fukuma et al. 1998). However, the initial studies supporting this hypothesis were conducted in Japanese patients. When extended to European patients the relationship between ISDR mutations and interferon response is less clear, as the number of 'mutant' ISDR sequences isolated from European patients is small (Rispetter et al. 1998). A recent study showed no association between ISDR sequences and response to interferon in American patients (Chung et al. 1999).

1.7 The quasispecies nature of HCV

Within an infected individual HCV exists as a swarm of distinct, but closely related, viral genomes, referred to as quasispecies (Martell et al. 1992; Farci et al. 1997; Gomez et al. 1999). This population of viral genomes behaves as if it were a single species (hence the prefix 'quasi') (Domingo et al. 1995). The concept of quasispecies was first suggested by Eigen, (1971) and provides a model for rapid evolution of a biological system under selective pressure. It may be the quasispecies nature of HCV that gave rise to the genotypes and subtypes of HCV that we see today.

1.7.1 Quasispecies and persistent infection

The existence of HCV as a quasispecies has several biological implications, the most important of these being its role in the establishment of persistent infections. It is possible that the host immune response is unable to contain such a wide spectrum of viral genomes (Farci et al. 1997). As neutralising antibody responses are directed at a portion of the E2 gene, the HVR-1, immune pressure is likely to be the main factor that drives the evolution in this region of the virus (Domingo, 1998a; Domingo et al. 1998b).

Quasispecies constituents may display different cellular tropisms. There is some evidence that HCV can replicate in cells other than those of the liver, such as peripheral blood mononucleocytes (PBMCs) (Sangar and Carroll, 1998). The quasispecies distributions in the liver and in PBMCs differ from each other and from those in the circulation (Cabot et al. 1997; Maggi et al. 1997; Navas et al. 1998). Differential tissue tropism suggests that different quasispecies constituents use different host cell receptors (Maggi et al. 1999). Such selectivity may allow the virus to infect cell types that are immunoprivileged, providing another possible mechanism by which HCV establishes persistent infection

(Gomez et al. 1999). However, experiments to detect HCV replication in PBMCs may be flawed, as they rely on the detection of negative sense RNA by PCR, which is prone to false-positive results when high levels of the corresponding positive sense RNA are present.

Some quasispecies constituents may be defective. Defective viral particles are not cytotoxic and can modulate the replication of wild-type virus, in the process reducing the extent of host cell death and facilitating viral persistence (Steinhauer and Holland, 1987; Martell et al. 1992; Higashi et al. 1993; Gomez et al. 1999).

1.7.2 Quasispecies and progression of liver disease

A more complex population of virus may be more likely to contain variants with propensity to cause hepatic damage. Alternatively, the broad immune response generated may be more cytotoxic when directed against certain constituents of the quasispecies, which are more apt to appear when the quasispecies diversity is wider. Several studies have demonstrated a relationship between quasispecies diversity and histological markers of liver damage (Honda et al. 1994; Gretch et al. 1996; Hayashi et al. 1997; Wyatt et al. 1998). However, other studies have failed to show such a correlation (Brambilla et al. 1998; Leone et al. 1998; Lopez-Labrador et al. 1999). It should be noted that these studies are usually concerned with the quasispecies distribution found in the circulation which may not reflect that in the liver (Cabot et al. 1997; Maggi et al. 1997). Since severity of liver disease is also associated with the duration of HCV infection, apparent association of quasispecies diversity with liver damage may reflect the accumulation of mutations over time.

1.7.3 Quasispecies and acute infection

Even less is known about HCV quasispecies evolution during acute infection. Studies in chimpanzees suggest that although an inoculum may contain a spectrum of HCV quasispecies, only certain constituents are transmitted during infection (Farci et al. 1997; Wyatt et al. 1998). The quasispecies distribution appears to be relatively homogeneous initially, becoming more diverse over time (Ni et al. 1997; Manzin et al. 1998). This broadening of the quasispecies diversity is probably a result of both the selective pressure applied by the host immune response and the evolutionary drift of the virus. Studies in the immunosuppressed such as bone marrow transplant recipients (Ni et al. 1999) and individuals with hypogammaglobulinaemia (Booth et al. 1998b) have shown that without immune pressure the quasispecies distribution is indeed homogeneous. However, this picture of acute infection with HCV, where very few species give rise to the complex quasispecies distribution seen later in chronic infection, is probably oversimplified. It is more likely that many minor species, not detected by current techniques, are also transmitted in the initial inoculum and that these species emerge as dominant species if environmental conditions are favourable.

It has also been observed that the level of quasispecies complexity seen in the early stages of infection is related to the outcome of the disease, whether the infection clears or whether chronic infection is established (Ray et al. 1999). However, this observation is difficult to substantiate since acute infection is very often sub-clinical, and in chimpanzees the proportion of infections that persist is much smaller than in humans (Bassett et al. 1998; 1999).

1.7.4 Quasispecies and the response to antiviral therapy

The diversity of HCV presents an obstacle to the design of broadly effective vaccines and antiviral drugs, as HCV possesses enormous potential for vaccine escape and drug resistance. Several studies have correlated high quasispecies diversity with failure to respond to treatment with interferon (Le Guen et al. 1997; Polyak et al. 1997; Toyoda et al. 1997; Pawlotsky et al. 1998b). Interferon itself appears to apply selective pressure to HCV in non-responders, as it has been observed that in these individuals the quasispecies distribution becomes more homogenous following treatment. While interferon treatment appears to facilitate immune clearance of the majority of quasispecies constituents in these patients, a few resistant strains remain (Nagasaka et al. 1996; Gonzalez-Peralta et al. 1997; Polyak et al. 1998).

1.8 HCV quasispecies and PCR studies

A convenient way to characterise HCV infecting an individual is to obtain the nucleotide sequence of a fragment of the genome amplified from a serum or plasma sample by the polymerase chain reaction (PCR). However, such sequences are only a "consensus" of all the sequences present in the quasispecies population. The consensus sequence may be identical to the sequence of the most abundant species, but if the dominant strain accounts for a minority of the total population then the consensus sequence may not reflect any single variant, as it describes the quasispecies as a whole. The components of a quasispecies can be studied separately. One way to do this is by cloning PCR products derived from a portion of the genome and obtaining the nucleotide sequence of each cloned species. However, such methods cannot provide a complete picture of a quasispecies population. Only relatively abundant variants are likely to be amplified, and mismatches in

the chosen primer sites may make amplification of certain variants impossible using a particular PCR protocol. The enzyme frequently used for amplification, *Taq* polymerase, can sometimes incorporate the wrong nucleotide during PCR, giving rise to virtual mutants that never existed in nature (Bracho et al. 1998). Conversely, when the viral load within a patient is low there is a greater risk of template resampling, resulting in the under estimation of viral diversity.

1.9 Quasispecies assessment by single-strand conformational polymorphism analysis

A fairly comprehensive picture of quasispecies distributions can be drawn from cloning and sequencing of PCR products. As nucleotide sequencing is a labour intensive procedure, several alternative methods have been used to estimate viral diversity. The most widely used of these in the study of HCV quasispecies is single strand conformational polymorphism (SSCP) analysis (Enomoto et al. 1994; Moribe et al. 1995; Pawlotsky et al. 1998b). SSCP can detect around 80% of single base mutations (Cotton, 1993). In this procedure PCR products derived from a mixed population, or from clones, are denatured and then electrophoresed through a non-denaturing polyacrylamide gel. During electrophoresis the DNA either reanneals to form double stranded DNA or remains single stranded and forms secondary structures that influence the migration of the single stranded DNA through the formation of different structures and hence alter the speed of migration through the gel matrix. Single stranded conformations are sequence dependent; mutations can result in the formation of different structures and hence alter the speed of migration through the gel. Mutations occurring in the 'loops' of a secondary structure will not affect the conformation. It is for this reason that SSCP is unable to detect all single base mutations.

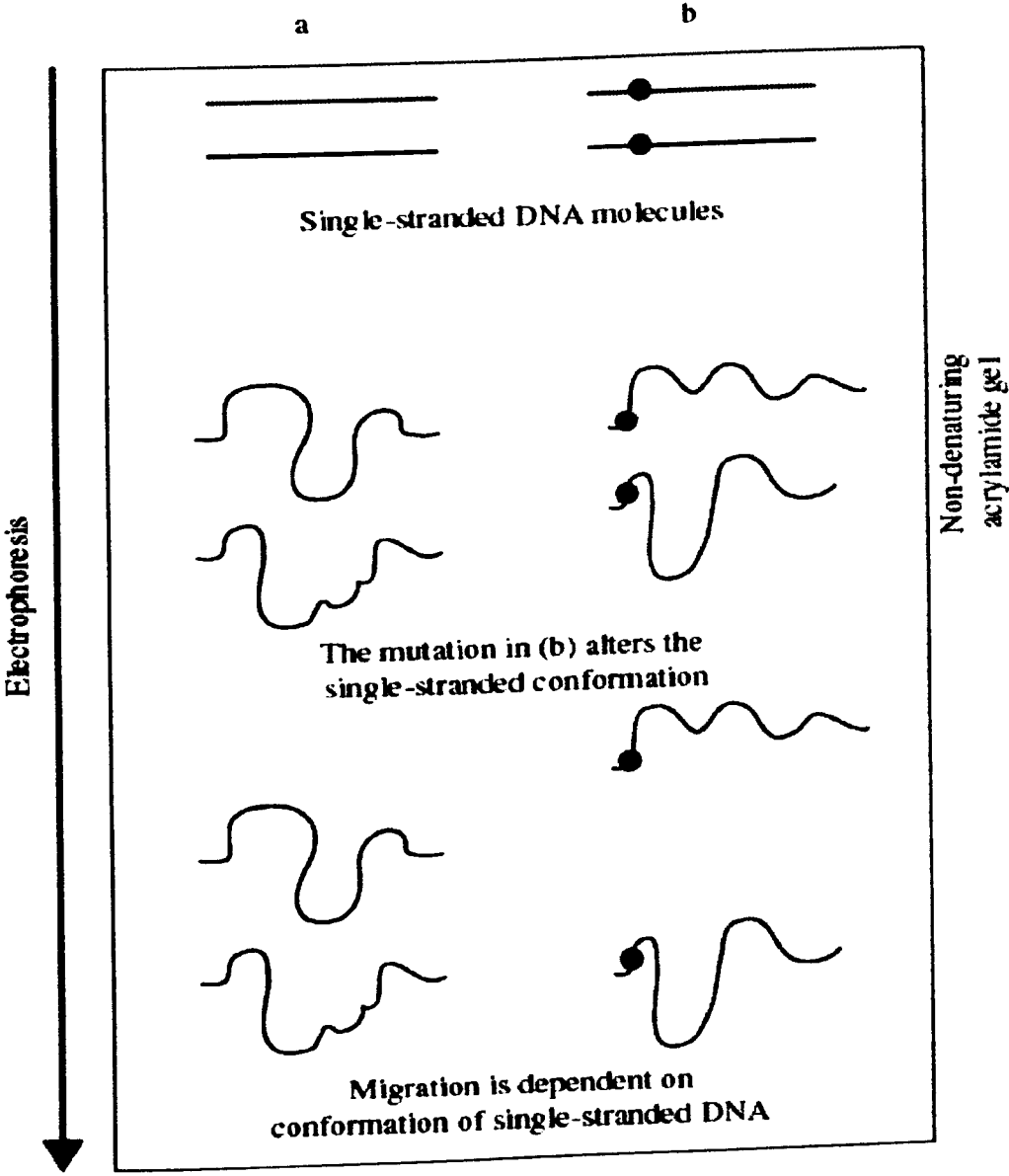


Figure 1.3: Schematic diagram to illustrate the principles of sequence mutation detection by SSCP analysis.

1.10 Quasispecies analysis using the heteroduplex mobility assay

Another mutation detection technique that has been applied to the analysis of HCV quasispecies is heteroduplex analysis, also known as the heteroduplex mobility assay (HMA) (Sullivan et al.1998). This method involves the formation of heteroduplexes between a ‘wild-type’ molecule and the test DNA. DNA containing mismatched base pairs

will be retarded in a non-denaturing gel matrix during electrophoresis in comparison to the homoduplex DNA. The distance between the homoduplexes and heteroduplexes is related to the number of mismatched base pairs. By using the PCR product derived from the dominant species of HCV as the tester, the changes in quasispecies populations over a period of time can be tracked.

The migration of heteroduplex molecules, particularly those with only one mismatched base pair, is not always different from that of the homoduplex molecules. The sensitivity of HMA is estimated to be 80% (Cotton, 1993). However, a combination of SSCP and HMA can achieve a mutation detection rate of almost 100% (Wang et al. 1998).

1.11 Quasispecies analysis using denaturing gradient gel electrophoresis

Denaturing gradient gel electrophoresis (DGGE) is another method for detecting mutations in DNA molecules. It has been widely used for the detection of mutations in human genes (Myers et al. 1987; Cremonesi et al. 1997; Guldberg et al. 1997) but rarely used to study viral genes (Woodward et al. 1994). DGGE can detect close to 100% of mutations in double stranded DNA (Cotton, 1993) and it is for this reason that I have selected this technique for the analysis of HCV quasispecies in this thesis. In the DGGE procedure, double stranded DNA, in the form of PCR product derived from clones, is electrophoresed through an acrylamide gel containing a gradient of denaturant that increases in the direction of electrophoresis. The double stranded DNA molecules will melt when they reach a part of the gel that is sufficiently denaturing. The DNA does not simply "unzip" when it melts. Instead, "bubbles" of single stranded DNA form at melting domains within the molecule. As conditions become more denaturing, more domains will melt and, eventually, single stranded regions retard the movement of double stranded DNA in the gel matrix. Sequence

differences of as little as one base can dramatically alter the stability of the melting domains. If the highest temperature melting domain is denatured then the DNA will be completely single stranded and the migration no longer dependent on sequence; electrophoresis should be stopped before this happens. In order to detect mutations along the whole length of a DNA molecule a "GC-clamp" must be attached to one of the PCR primers (Sheffield et al. 1989). The clamp sequence is extremely GC rich to prevent the complete denaturation of the double stranded DNA, hence allowing detection of virtually all mutations. The sensitivity of DGGE is highly dependent on the quality of the gradient gels and the differences in migratory positions of DNA molecules are not necessarily related to the number of nucleotide differences.

1.12 Aims of this study

The broad aim of this thesis was to investigate the molecular variability of HCV. Inter-host variation was studied using a PCR-RFLP assay, modified to allow rapid screening of a large number of samples. The DGGE technique was developed in order to screen for host quasispaces changes on a cross-sectional and longitudinal basis.

I then tested three hypotheses. 1: The distribution of HCV genotypes is higher in similar to that found in the rest of northern Europe; 2: the mode of acquisition of HCV is the genotype and complexity of the circulating virus is more likely to be detected by more than one HCV test.

HCV does not mutate faster than other RNA viruses.

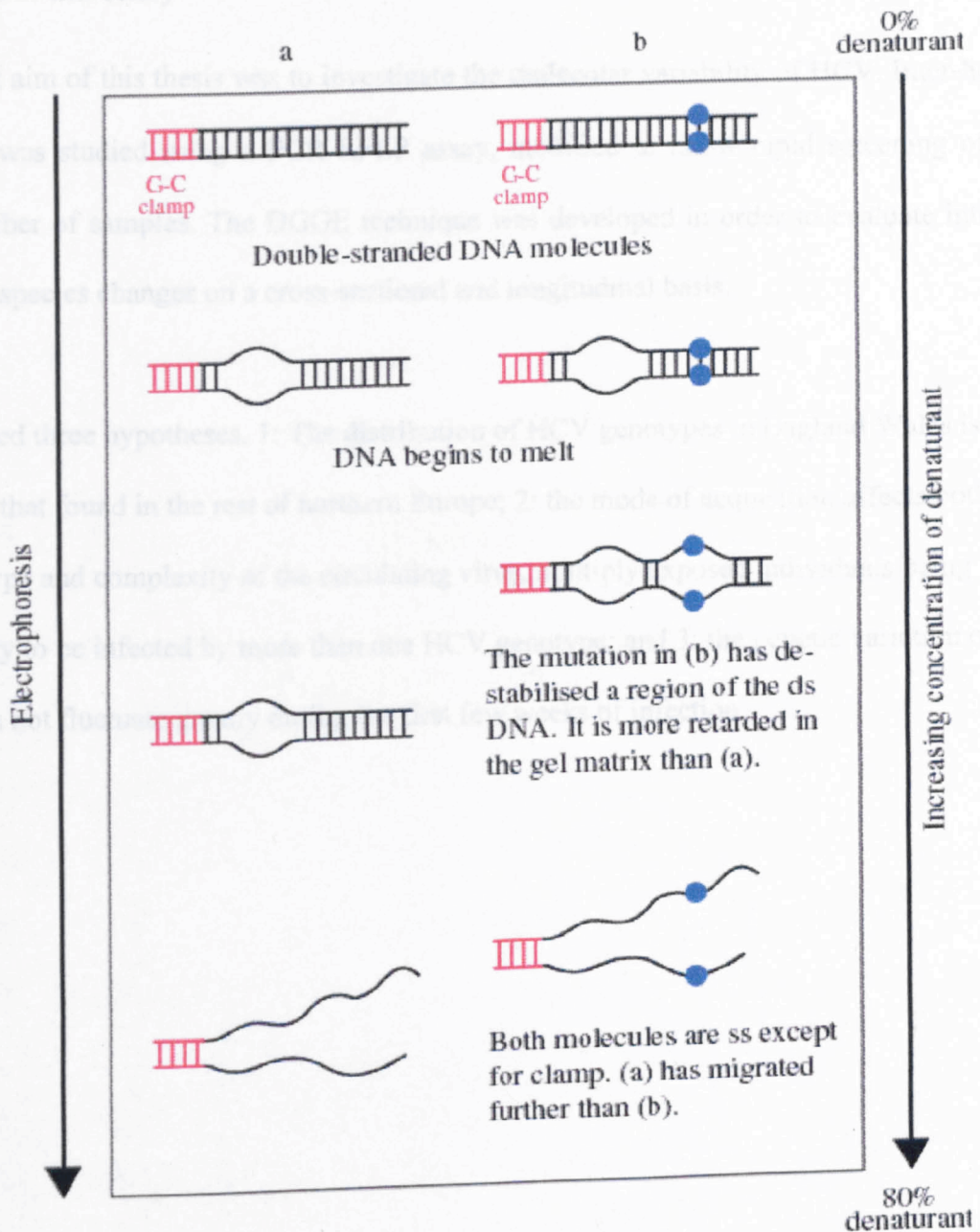


Figure 1.4: Schematic diagram to illustrate the principles of sequence mutation detection by DGGE.

1.12 Aims of this study

The broad aim of this thesis was to investigate the molecular variability of HCV. Inter-host variation was studied using a PCR-RFLP assay, modified to allow rapid screening of a

1.12 Aims of this study

The broad aim of this thesis was to investigate the molecular variability of HCV. Inter-host variation was studied using a PCR-RFLP assay, modified to allow rapid screening of a large number of samples. The DGGE technique was developed in order to evaluate intra-host quasispecies changes on a cross-sectional and longitudinal basis.

I then tested three hypotheses. 1: The distribution of HCV genotypes in England Wales is similar to that found in the rest of northern Europe; 2: the mode of acquisition affects both the genotype and complexity of the circulating virus, multiply exposed individuals being more likely to be infected by more than one HCV genotype; and 3: the genetic variation of HCV does not fluctuate greatly during the first few weeks of infection.

Chapter 2

Materials and Methods

2.1 Amplification of HCV 5' Non-Coding Region (NCR)

2.1.1 RNA Extraction

RNA was extracted from 100 microlitres (μ l) of sera or plasma using the Amplicor HCV Specimen Preparation Kit (Roche Diagnostic Systems, Branchburg USA). The final pellet was resuspended in 50 μ l of nuclease free water.

2.1.2 Synthesis of complementary DNA (cDNA)

Extracted RNA (22.2 μ l) was reverse transcribed in the following reaction: 1X PCR buffer (Life Technologies, Paisley, U.K.), 5 mM $MgCl_2$ (Life Technologies), 1 mM each of the four deoxynucleotide triphosphates (dNTPs) (Life Technologies), 3.3 μ M random hexamers (Pharmacia Biotech U.K., St. Albans, Herts.), 0.34 units RNasin (Promega U.K., Southampton, U.K.), 5 units of Murine Maloney Leukemia Virus (MMLV) reverse transcriptase (Life Technologies) and nuclease free water to make the final reaction volume 40 μ l.

2.1.3 Nested Polymerase chain reaction (PCR)

Primary amplification was carried out in the following reaction; 1X PCR buffer, 2 mM $MgCl_2$, 5 pmoles of sense primer 57, 5 pmoles of antisense primer 321, 0.625 units of Taq polymerase (Life Technologies), 10 μ l of cDNA solution and nuclease free water to make the final reaction volume 50 μ l. Reactions were heated to 94°C for 30 s followed by 35 cycles of: 94°C for 30 s, 62°C for 40 s and 72°C for 50 s.

Secondary amplification was carried out in the following reaction: 1X PCR buffer, 2 mM MgCl₂, 1 mM of each of the four dNTPs, 25 pmoles of sense primer 126 and 25 pmoles of antisense primer 299, 0.625 units of Taq polymerase and 2 µl of primary PCR product. Cycling conditions were 25 cycles of 94°C for 30 s, 68°C for 40 s and 72°C for 30 s.

Primer sequences were those described by (Lin et al. 1992);

Sense 57: 5' AGCGTCTAGCCATGGCGT

Antisense 321: 5' GCACGGTCTACGAGACCT

Sense 126: 5' GTGGTCTGCGGAACCGG

Antisense 299: 5' GGGCACTCGCAAGCACCC

The inner primer pair amplified a 174-bp fragment between positions -199 and -26 of the HCV genome, numbered according to (Choo et al. 1991).

2.1.4 Agarose gel electrophoresis

PCR products were electrophoresed through 2% agarose (SB fine gel, Severn Biotech Ltd., Kidderminster, Worcs. U.K.) in 1X TRIS-borate-EDTA (TBE) buffer (supplied as 10X stock by Life Technologies). 1-kilobase DNA markers (Life Technologies) were run alongside the PCR products to assess their size. Gels were stained in ethidium bromide solution at a concentration of 5 µg per ml and visualised by ultraviolet transillumination.

2.2 Genotyping of HCV using the 5'NCR

2.2.1 Restriction digests of second round PCR product

The 174 base-pair second round PCR product, amplified from the HCV 5'NCR as described in section 2.1, was digested with the restriction enzymes *Sac*II, *Mva*I, *Hin*II and

*Bst*UI (Pohjanpelto et al. 1996) in four individual 10 µl reactions comprising 0.5 µl of enzyme, 1 µl reaction buffer (supplied with enzyme), 4.5 µl of PCR product and 4 µl of sterile water. Following incubation at 37°C for 1.5 hr, 60°C for *Bst*UI, digested samples were heated to 80°C for 15 min to denature the enzymes. *Scr*FI, *Hin*FI and *Bst*UI were supplied by New England Biolabs (Hitchin, Herts. U.K.) and *Mva*I by Roche Molecular Biochemicals, (Lewes, East Sussex, U.K.).

2.2.2 Polyacrylamide gel electrophoresis (PAGE)

Digests were electrophoresed through an MDE™ gel (Flowgen, Lichfield, Staffordshire, U.K.), prepared according to the manufacturers instructions in 0.6X TBE. *Scr*FI and *Mva*I digests were loaded into one well and *Hin*FI and *Bst*UI digests into the contiguous well. PhiX174 DNA/Hinf I markers (Promega U.K.) were loaded alongside the digests to assess the RFLP fragment sizes. A 48-well shark tooth comb was used for loading. The gels were run in 0.6X TBE at 170 V for 3 hrs then stained with SYBR Green I (Flowgen, Lichfield, Staffordshire, U.K.), at a concentration of 10 µl of stock solution per 100 ml of 0.6X TAE buffer (supplied as 10X stock by Life Technologies), and visualised by ultraviolet transillumination.

2.2.3 Sequencing

Samples whose genotypes could not be determined by PCR-RFLP were sequenced with primers 126 and 299. Second round PCR products were purified by electrophoresis through 2% agarose followed by DNA recovery using the Igenie DNA extraction kit (Immunogen, Sunderland, Tyne and Wear, U.K.). Sequencing reactions were performed using the ABI Prism DNA sequencing kit (PE Applied Biosciences, Warrington, U.K.) and reactions were electrophoresed using the ABI 373 automated sequencer (PE Applied Biosciences).

Sequences were aligned and phylogenetically analysed with other HCV 5'NCR sequences by the CLUSTAL algorithm in the MEGALIGN programme of the LASERGENE system (DNASTAR Inc, Madison, Wisconsin, USA).

2.2.4 Cloning

PCR products whose RFLP profile indicated the presence of more than one genotype, or those that could not be assigned a genotype, were cloned using the TOPO TA cloning kit (Invitrogen BV, De Schelp, The Netherlands). Colonies with inserts were picked and amplified directly using PCR conditions described earlier for second round HCV 5'NCR PCR. The products then underwent either RFLP analysis as described in section 2.2.1 or DNA sequencing as described in section 2.2.3.

2.3 Statistical analysis

Single variable analyses (Everitt, 1997) were performed to investigate whether the genotype distributions differed for each patient group by age or sex. The genotypes were grouped into 1a, 1b, 2, 3a, 3b, 4, 5 and 6. Samples identified as containing genotype 1 for which a subtype could not be assigned (1x) and those with mixed genotypes were excluded from this analysis, since the numbers of such samples was small. Log-linear regression analysis was performed in order to investigate any associations between genotype distribution and age, sex or patient group.

2.4 Amplification of HCV Non-Structural Region 5b (NS5b)

2.4.1 RNA extraction and cDNA synthesis

RNA extraction and cDNA synthesis used the same procedures as those described in 2.1.1 and 2.1.2.

2.4.2 Single round PCR

Amplification was carried out in the following reaction: 1X Opti-prime™ PCR buffer #4 (Stratagene, Cambridge, U.K.), 25 pmoles of sense primer 1204, 25 pmoles of antisense primer 1203, 0.625 units of Taq polymerase, 10 µl of cDNA solution and nuclease free water to make the final reaction volume 50 µl. Reactions were heated to 94°C for 30 seconds followed by 35 cycles of: 94°C for 30 s, 54°C for 40 s and 72°C for 50 s, followed by 30 s at 72°C.

Primer sequences were as follows (Mellor et al. 1995):

Sense primer 1204; 5'GGAGGGGCGGAATACCTGGTCATAGCCTCCGTGAA

Antisense primer 1203; 5'ATGGGGTTCTCGTATGATACCCGCTGCTTTGACTC

These sequences amplified a 401-bp fragment between positions 7903 and 8309 of the HCV genome (numbered according to (Choo et al. 1991), which was detected by agarose gel electrophoresis as described in 2.1.4.

2.5 Genotyping of HCV using the NS5b region

2.5.1 Gel purification of PCR products

PCR products were purified by electrophoresis through 2% agarose followed by recovery of DNA using the Igenie DNA extraction kit (Immunogen, Sunderland, Tyne and Wear, UK).

2.5.2 Cloning and sequencing of purified PCR products

Purified NS5b PCR products were cloned using the TOPO TA cloning kit. Colonies containing inserts were amplified directly using the PCR conditions described in 2.5.2. Colony PCR products were gel purified as described in 2.6.1 and sequenced with primers 1204 and 1203. Sequencing reactions were performed using the ABI Prism DNA sequencing kit and reactions were run on the ABI 373 automated sequencer. A 222-bp fragment of the NS5b sequence was aligned and phylogenetically analysed with other sequences from this region obtained from the Genbank database. The CLUSTAL V algorithm in the MEGALIGN program within the LASERGENE package (DNASTAR Inc, Madison, Wisconsin, USA).

2.6 Analysis of HCV quasispecies by denaturing gradient gel electrophoresis (DGGE)

Analysis of HCV quasispecies infecting individuals was performed using DGGE analysis and DNA sequencing of cloned PCR products. Single-round PCR was performed to amplify fragments of the 5'NCR, NS5b and HVR1, in order to minimise the number of PCR cycles and hence Taq misincorporation errors.

2.6.1 Single round HCV 5'NCR PCR

Amplification was carried out in the following reaction; 1X PCR buffer (supplied as part of EXPAND high fidelity PCR kit by Roche Molecular Biochemicals, Lewes, East Sussex, U.K.), 1.5 mM MgCl₂ (supplied as part of EXPAND high fidelity PCR kit by Roche Molecular Biochemicals), 1 mM each of the four dNTPs, 25 pmoles of sense primer 126, 25 pmoles of antisense primer 299, 0.7 units of EXPAND high fidelity polymerase (Roche Molecular Biochemicals), 10 µl of cDNA solution and nuclease free water to make the final reaction volume 50 µl. Reactions were heated to 94°C for 30 s followed by 35 cycles of: 94°C for 30 s, 65°C for 40 s and 72°C for 50 s, followed by 30 s at 72°C.

2.6.2 Single round HCV NS5b PCR

Amplification was carried out as described in 2.4.2, except that 0.7 units of EXPAND high fidelity polymerase were used in place of Taq polymerase to reduce the number of artefactual mutations due to polymerase incorporation errors.

2.6.3 Gel purification of HCV NS5b PCR products

PCR products were purified by electrophoresis through 2% agarose followed by recovery of DNA using the Igenie DNA extraction kit.

2.6.4 Amplification of HCV envelope region, incorporating HVR-1

RNA extraction and cDNA synthesis used the same procedures as those described in sections 2.1.1 and 2.1.2. Primary amplification was carried out in the following reaction: 1X EXPAND PCR buffer, 2 mM MgCl₂, 5 pmoles of sense primer HVR-S1, 5 pmoles of antisense primer HVR-A1, 0.7 units of EXPAND high fidelity polymerase, 10 µl of cDNA solution and nuclease free water to make the final reaction volume 50 µl. Reactions were

heated to 94°C for 1 min followed by 30 cycles of: 94°C for 1 min, 56°C for 1 min and 72°C for 1 min.

Secondary amplification was carried out in the following reaction: 1X PCR buffer, 1.5 mM MgCl₂, 1 mM of each of the four dNTPs, 25 pmoles of sense primer HVR-S2 and 25 pmoles of antisense primer HVR-A2, 0.7 units of EXPAND high fidelity polymerase and 2 µl of primary PCR product. Cycling conditions were 15 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1 min.

Primer sequences were as follows (Toyoda et al. 1998):

HVR-S1: 5' TGGGACACATGATGATGAACTGGT

HVR-A1: 5' CGGTGCTGTTTATGTGCCAACTGCC

HVR-S2: 5' TACTACTCCATGGTGGGAGACTGGGC

HVR-A2: 5' GATGTGCCAGCTGCCATTGG

These sequences amplify a 189 bp nested PCR product between positions 1082 and 1271 of the HCV genome (numbered according to (Choo et al. 1991). This fragment incorporates HVR-1 and was detected by agarose gel electrophoresis as described in 2.1.4.

2.6.5 Cloning

HCV 5'NCR single-round PCR products, HCV envelope/HVR1 single round products and purified HCV NS5b single-round PCR products were cloned using the TOPO TA cloning kit. Colonies with inserts were picked and amplified directly using PCR conditions described below in Sections 2.6.6 to 2.6.8.

2.6.6 Colony PCR to amplify 5'NCR inserts and the attachment of a GC-clamp

Colony PCR conditions were identical to the PCR conditions used for second round amplification of the HCV 5'NCR (2.1.3) with the exception of the number of cycles, which was reduced to 20 and the following forward primer sequence to which a 40 bp GC-clamp was attached (Sheffield et al. 1989):

Sense126 (clamp): 5' CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCC
CGCCCGGTGGTCTGCGGAACCGG

The 214-bp PCR product was visualised by agarose gel electrophoresis (2.1.4).

2.6.7 Colony PCR to amplify NS5b inserts and the attachment of a GC-clamp

Colony PCR conditions were identical to the PCR conditions used for single round amplification of the HCV NS5b region (2.6.2) with the exception of the number of cycles, which was reduced to 20 and the following forward primer sequence to which a 40 base pair GC-clamp was attached (Sheffield et al. 1989):

Sense primer 1204 (clamp): 5'CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGC
CCCCGCCCCGGGAGGGGCGGAATACCTGGTCATAGCCTCCGTGAA

The 441-bp product was visualised by agarose gel electrophoresis (2.1.4).

2.6.8 Colony PCR to amplify envelope/HVR1 inserts and the attachment of a GC-clamp

Colony PCR conditions were identical to the PCR conditions used for second round amplification of the HCV envelope/HVR1 region (2.6.4) with the exception of the following forward primer sequence to which a 40 base pair GC-clamp was attached (Sheffield et al. 1989):

Sense primer HVR-S1 (clamp): 5'CGCCCGCCGCGCCCCGCGCCCGTCCCG
CCGCCCCCGCCCGTGGGACACATGATGATGAACTGGT

The 229-bp product was visualised by agarose gel electrophoresis (2.1.4).

2.6.9 Preparation of a denaturing gradient polyacrylamide gel for simultaneous analysis of 5'NCR and NS5b colony PCR products

Acrylamide solutions were prepared as two stock solutions. One solution contained 30% denaturants and was prepared as follows in distilled water: 12% polyacrylamide (37.5:1) (supplied as Protogel by National Diagnostics, Hull, Yorks. U.K.); 0.6X TRIS-acetate EDTA (TAE) buffer; 12% formamide (BDH, Poole, Dorset, U.K.) and 2.1 M urea (BDH). The second solution contained 70% denaturants and was prepared as follows in distilled water: 12% polyacrylamide (37.5:1); 0.6X TAE buffer; 28% formamide and 4.9 M urea.

The gradient gel was poured using 30 ml of each gel solution and a gravity driven gradient maker (GRI, Braintree, Essex U.K.). The polymerisation reaction was initiated with 13 μ l TEMED (Life Technologies) per 30 ml of gel solution and catalysed by 250 μ l of 10% ammonium persulphate (BDH) per 30 ml of gel solution. Wells were formed with a 48 well square-toothed comb. The gel was allowed to polymerise for 1 hr.

2.6.10 Preparation of a denaturing gradient polyacrylamide gel for analysis of envelope/HVR-1 colony PCR products

Acrylamide solutions were prepared as two stock solutions. One solution contained 10% denaturants and was prepared as follows in distilled water: 12% polyacrylamide (37.5:1); 0.6X TAE buffer; 4% formamide and 0.7 M urea. The second solution contained 65% denaturants and was prepared as follows in distilled water: 12% polyacrylamide (37.5:1); 0.6X TAE buffer; 26% formamide and 4.55 M urea. The gradient gel was poured as described in Section 2.6.9.

2.6.11 Loading and electrophoresis

Clamped PCR products (3-5 μ l of each), generated as described in 2.6.4 and 2.6.5, were mixed with 2-4 μ l of loading buffer (10% Ficoll-400; 10mM TrisHCl (pH 7.5); 50mM EDTA; 0.25% xylene cyanol FF) and applied to the gel. When two regions were analysed simultaneously colony PCR products were pooled prior to loading the gel. Electrophoresis was in 0.6X TAE buffer at 60°C and at 100 V for 16 hr, two gels running simultaneously. Running buffer was continuously circulated to maintain constant temperature throughout the electrophoresis tank.

2.6.12 Gel staining

The gel was stained with SYBR Green I (Flowgen) at a concentration of 10 μ l stock solution per 100 ml of 0.6X TAE buffer (supplied as 10X stock by Life Technologies). Bands were visualised by ultraviolet transillumination.

2.6.13 Sequencing of variants

All colony PCR products with a unique migration position following DGGE were sequenced. The 5'NCR and envelope/HVR-1 products were sequenced as described in 2.2.3. The NS5b products were prepared for sequencing in the same way but reactions were carried out using the ABI Prism DNA sequencing kit with Big Dye™ terminators (PE Applied Biosciences) and reactions were electrophoresed using the ABI 377 automated sequencer. This alternative sequencing procedure ensured that the quality of data was sufficient to achieve the full-length sequence in both forward and reverse directions for this 441-bp PCR product.

2.6.14 Analysis of sequence data

Sequences were aligned using the CLUSTAL V algorithm in the MEGALIGN program from the LASERGENE package. Amino acid sequences were determined using the MEGALIGN program and were once again aligned using the CLUSTAL V algorithm within this program. Ratios of synonymous to non-synonymous mutations (ds/dn) were calculated for each alignment by dividing the number of synonymous changes by the number of non-synonymous changes. Sequences with large deletions were excluded from this analysis and single base deletions were not considered as sites of nucleotide change. Transition to transversion ratios were calculated by dividing the number of transitions by the number of transversions. Phylogenetic analysis was carried out using the CLUSTAL V algorithm in the MEGALIGN program. An additional method was also used to analyse the same sequences. DNADIST, from the PHYLIP suite of programs (Felsenstein, 1993), was used to generate a distance matrix and FITCH, from the PHYLIP suite of programs was used to relate the sequences to each other. TREEVIEW was used to display the output from FITCH as a dendrogram. Both CLUSTAL V and FITCH carry out pair-wise comparisons of sequences using a neighbour joining method. Input order of sequences into DNADIST was randomised 10 times for each tree produced with this method.

2.6.15 Calculating genetic complexity and diversity

Shannon entropy (S_n) values were calculated as a measure of genetic complexity:

$$S_n = [\sum_i (p_i \ln p_i)] / \ln N,$$

where N = total number of sequences and p_i = frequency of each sequence in the viral quasispecies.

Genetic diversity was calculated as the average genetic distance between species, using the CLUSTAL V algorithm in the MEGALIGN program.

2.6.16 Statistical analysis

Sn values were used to test the hypothesis that diversity of the 5'NCR differed from the diversity of the NS5b region. The data from each region were verified to be normal following application of the Shapiro-Wilks test. Sn values from each region were tested for equality using t-tests on mean values and variance ratio tests on standard deviations. The statistical significance of differences in Shannon entropy and genetic distance values between the three patient groups was tested with the Krustal–Wallis test.

Chapter 3

Distribution of HCV Genotypes in England and Wales.

3.1 Introduction

The hepatitis C virus (HCV) genome displays sequence diversity. Variants can be classified into six major genotypes, each of which contains a number of closely related subtypes (Simmonds et al. 1993; Simmonds, 1995; Smith and Simmonds, 1997). Different genotypes may carry different clinical significance (Dusheiko et al. 1994a; Dusheiko and Simmonds, 1994b; Kiyosawa, 1997; Lopez-Labrador et al. 1997; Bruno et al. 1997; Mihm et al. 1997). They may also influence response to interferon-alpha, currently the only therapeutic drug licensed in the treatment of HCV infection (Dusheiko and Simmonds, 1994b; Zein et al. 1996; Dusheiko et al. 1996; Kiyosawa, 1997; Isaacson et al. 1997).

HCV genotype distributions vary geographically (see introductory discussion in Section 1.6.2). Within a given geographical region different genotypes may predominate in different subpopulations (Pawlotsky et al. 1995; Basaras et al. 1997). Such subpopulations would include groups of people at high risk for acquiring HCV infection, e.g., haemophilia patients or IDUs. Immigrants can also bring with them HCV genotypes peculiar to their country of origin.

In England and Wales, the extent to which HCV genotypes differ in at-risk groups has not been investigated. Prospective capture of this information is required to determine what the current range of genotypes is. This will allow more informed management of infection (Dusheiko et al. 1996) and facilitate earlier recognition of outbreaks caused by new or unusual genotypes (Smith and Simmonds, 1997). I conducted a systematic study of prevailing genotypes in England and Wales and compared the results with international data.

3.2 Materials and Methods

3.2.1 Patients

Between August 1996 and March 1998 samples from 567 HCV RNA-positive individuals, referred to the PHLS hepatitis and retrovirus laboratory from 46 centres, were genotyped. Samples were from 77 haemophilia patients, 68 screened blood donors and 78 IDUs. Other samples were from 62 antenatal clinic (ANC) patients, and 86 genito-urinary medicine (GUM) clinic patients identified in the Unlinked Anonymous Seroprevalence Monitoring Programme (Unlinked Anonymous HIV Surveys Steering Group, 1996). The remaining samples were from 194 individuals investigated for liver disease. The six patient groups were analysed by sex and age, <40 yr and ≥ 40 yr, with the exception of the GUM group, where a cut-off of 35 yr was used.

3.2.2 Methods

A fragment of the 5'NCR of HCV infecting the 567 individuals in this study was amplified by nested RT-PCR and genotyped by the RFLP assay described in Sections 2.1 and 2.2. Amplicons that could not be assigned a genotype using this technique were subjected to direct DNA sequencing. Mixed genotypes were confirmed by cloning and RFLP. The genotype distribution data were analysed using statistical methods (Everitt B.S.,1997). These methods are described in full in Sections 2.2.3, 2.2.4 and 2.2.5.

Figure 3.1 is a schematic diagram showing the positions of the restriction sites for each of the genotypes and subtypes that this method can identify.

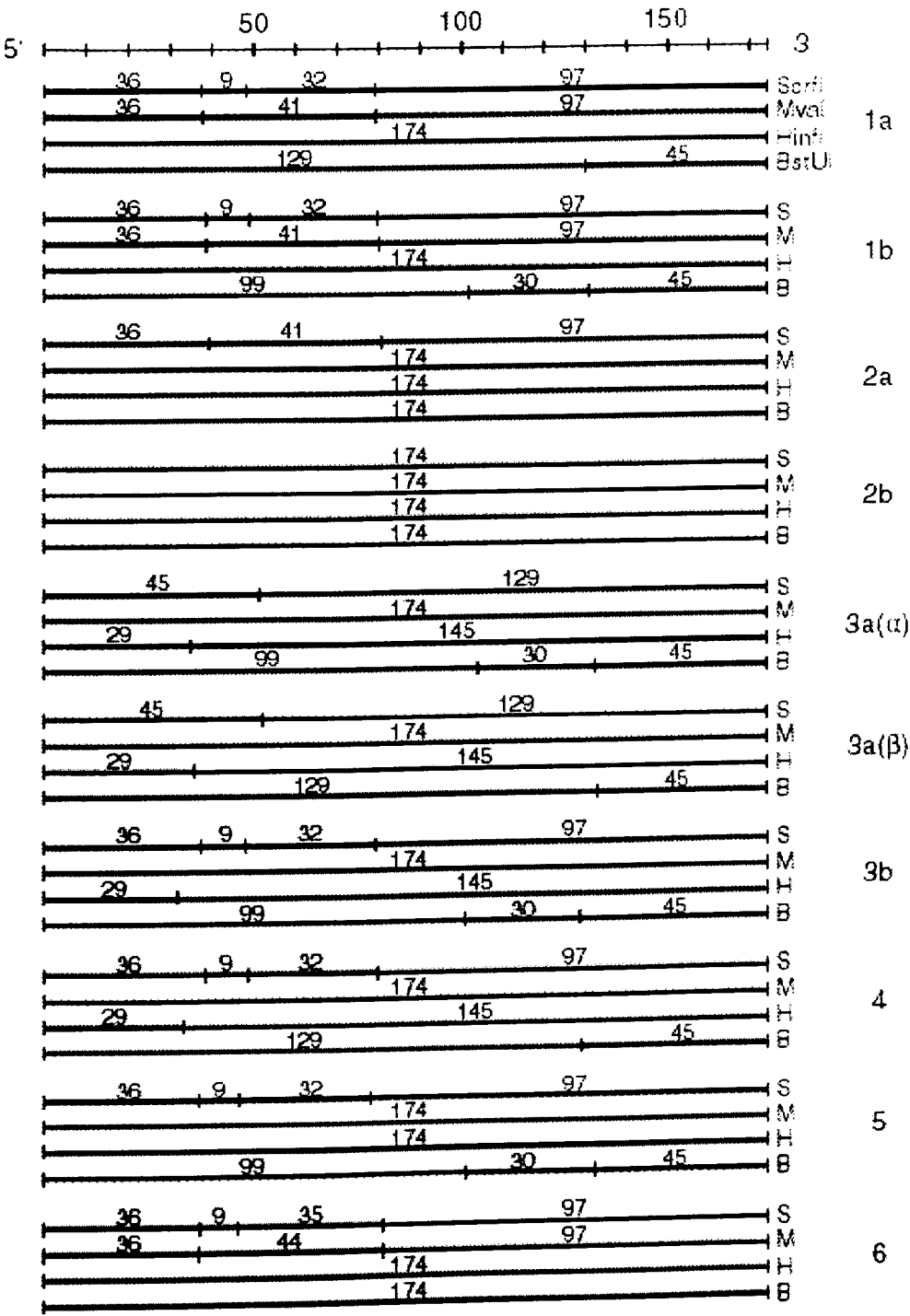


Figure 3.1: Restriction fragment length polymorphisms of the 174 base-pair fragment of the 5' non-coding region of the HCV genome used to distinguish genotypes and subtypes.

This procedure was developed from that described by (Pohjanpelto et al. 1996). Amplicons were digested as described in Section 2.2.1. In the early stages of developing this technique agarose was used as the gel matrix for electrophoresis. Each of the four digests were loaded

into a separate well and electrophoresed through a 4% Metaphor™ for 3-4 hr at 100 V. Gels were stained with ethidium bromide as described in Section 2.1.4 and visualised by UV transillumination. Figure 3.2 is an example of such a gel. Fragment sizes smaller than 97 base-pairs could not be visualised by this method. The larger fragments can also be hard to see if the PCR product is weak.

The first stage in the development of the procedure was the electrophoresis of digests through MDE™ gels (Flowgen). This offered two main advantages: digests could be pooled prior to gel loading, as described in Section 2.2.2, allowing a greater number of samples to be analysed simultaneously; and the electrophoresis time was shorter. Ethidium bromide could still be used to stain these gels as described in Section 2.1.4, but this did not allow visualisation of the smaller bands (those less than 50 bp). Staining the gels with SYBR-Green I, as described in Section 2.2.2, made the interpretation of ambiguous band patterns easier as the smaller bands became clearly visible. Figure 3.3 shows a typical genotyping gel after these modifications to the procedure had been made.

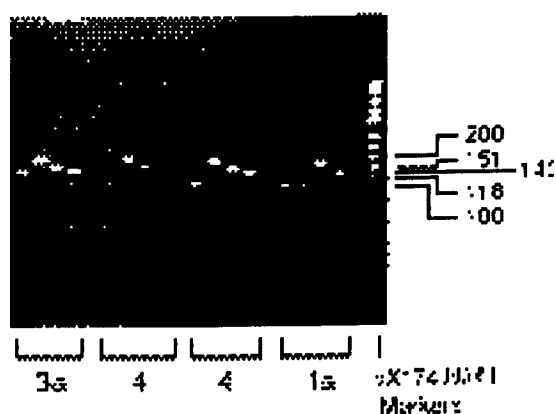


Figure 3.2: Restriction digests were electrophoresed through a 4% Metaphor™ gel and stained with ethidium bromide. Each of the four digests for each sample were loaded in a separate well in the following order, from left to right, ScaI, MvaI, HinfI and BstUI.

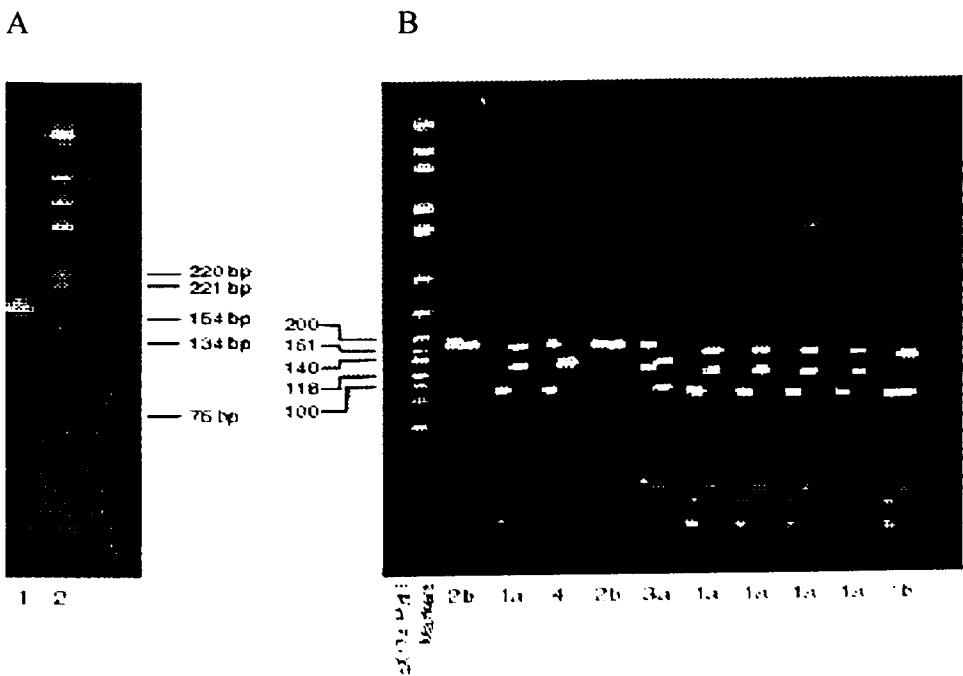


Figure 3.3: (a) The undigested 174 base pair nested PCR product amplified from the 5'NCR of HCV and visualised by agarose gel electrophoresis. (b) A typical genotyping gel. Digested PCR products were electrophoresed through a MDE™ gel which was stained with SYBR-green I. For each sample *Scrfl* and *MvaI* digests were loaded into the first well and *HinfI* and *BstUI* digests into the next well.

3.3 Results

3.3.1 Genotype distribution

The high-throughput PCR-RFLP procedure described in Sections 2.1 and 2.2 was used to investigate the distribution of HCV genotypes infecting 567 individuals. This data is shown

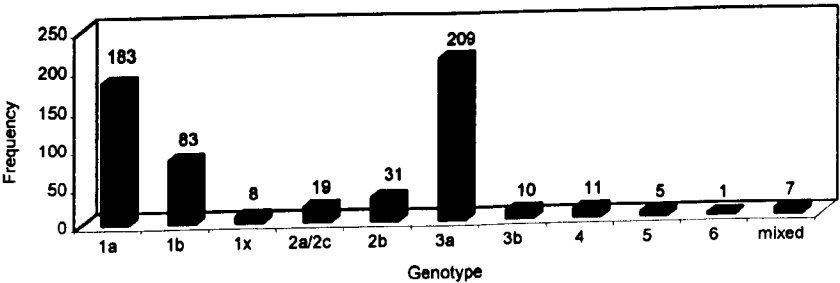


Figure 3.4: Distribution of HCV genotypes in 567 individuals determined by PCR-RFLP between August 1996 and March 1998.

Genotype distributions specific for haemophiliacs, blood donors, IDU, those attending ANCs, those attending GUM clinics and liver disease patients are displayed in Figures 3.5 - 3.10.

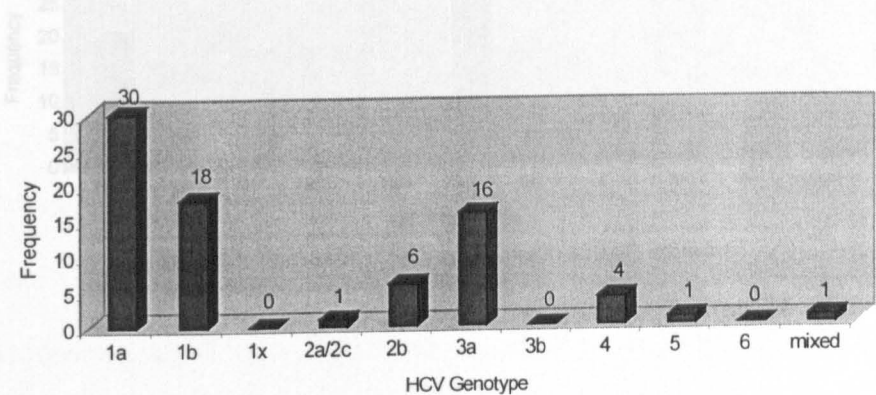


Figure 3.5: Distribution of HCV genotypes in 77 haemophilia patients

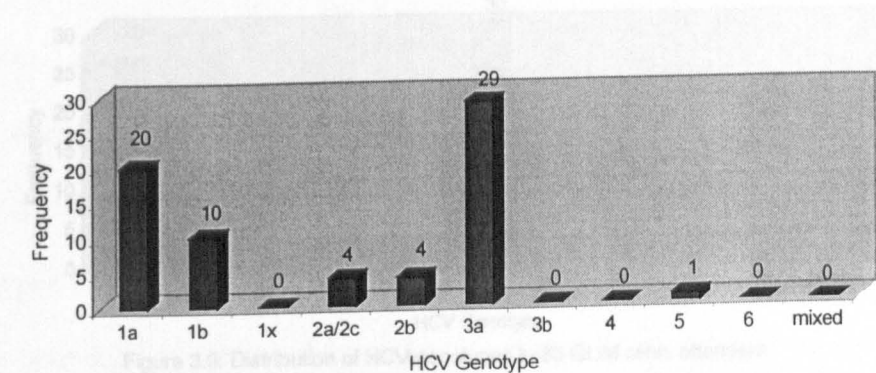


Figure 3.6: Distribution of HCV genotypes in 68 blood donors

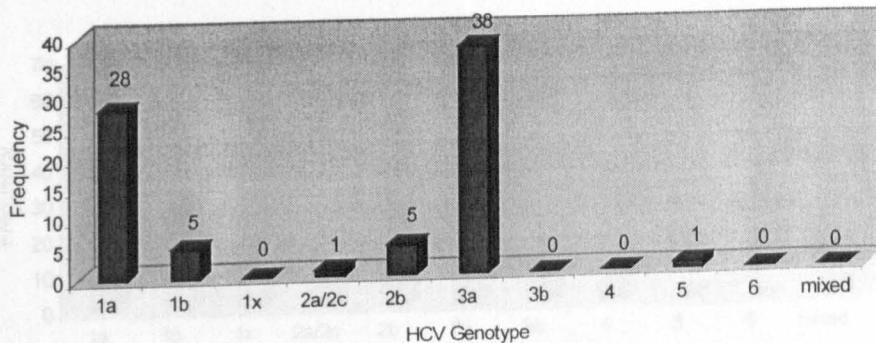


Figure 3.7: Distribution of HCV genotypes in 78 intravenous drug users

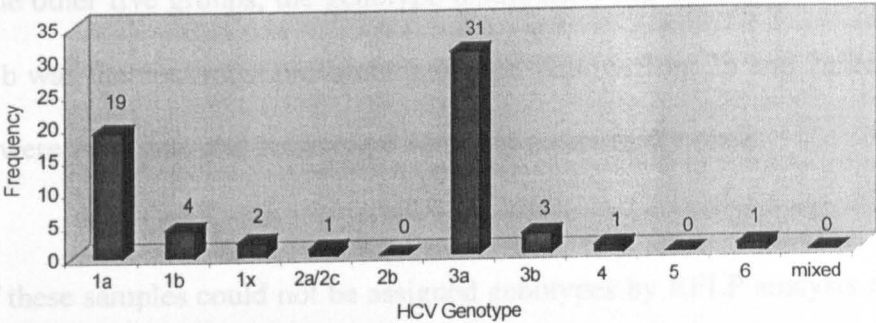


Figure 3.8: Distribution of HCV genotypes in 62 ANC attenders

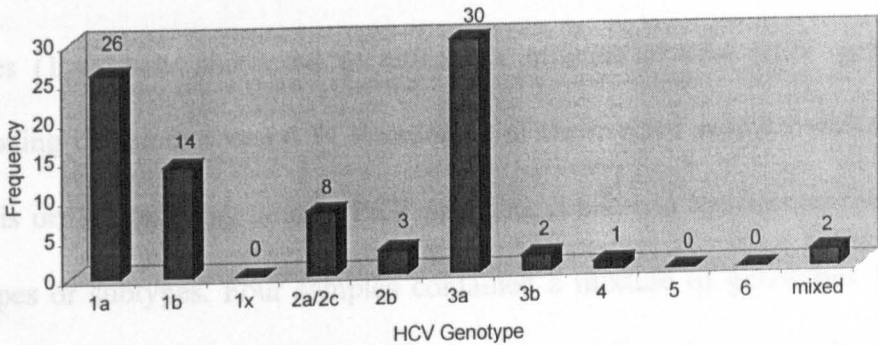


Figure 3.9: Distribution of HCV genotypes in 86 GUM clinic attenders

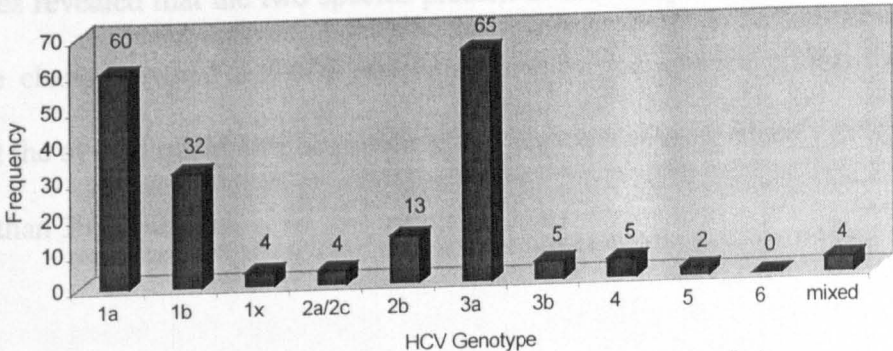


Figure 3.10: Distribution of HCV genotypes in 194 liver disease patients

Among haemophilia patients, the most common genotypes were 1a (39%), 1b (23%) and 3a (20%). In the other five groups, the genotype distribution was similar: 1a and 3a were co-prevalent, 1b was the next most prevalent genotype, followed by 2b and 2a/2c. Genotypes 3b, 4 and 5 were very rare, and genotype 6 was encountered only once.

Fourteen of these samples could not be assigned genotypes by RFLP analysis alone, so the second round PCR products of these samples were subjected to direct DNA sequencing. Following sequencing eight individuals were found to be infected with HCV genotype 1 but an existing subtype could not be assigned (represented as type 1x in Figures 3.4 – 3.10).

Eight samples (1%) were suspected to contain a mixture of two HCV genotypes or subtypes. Cloning the second round PCR products of these eight samples and performing RFLP analysis on the resulting colony PCR products confirmed the presence of multiple HCV genotypes or subtypes. Four samples contained a mixture of genotypes 1a and 1b, two samples contained a mixture of types 1a, 1b and type 1x, and one sample contained a mixture of genotypes 1b and 2b. The final sample appeared to contain a mixture of genotypes 3b and 4 following RFLP analysis of clones. However, subsequent sequencing of the clones revealed that the two species present in this sample were both genotype 4. A single base change created a *Bst*UI restriction site in one species giving the 3b RFLP pattern, but the overall nucleotide sequence in this region was more closely related to type 4 sequences than 3b sequences.

3.3.2 Statistical analysis

The hypothesis that genotype is associated with sex and age of patients was tested by Chi-square analysis (Everitt, 1997). Very few patients were of genotype 3b-6 and they were therefore excluded from this preliminary analysis. It was not appropriate to test for an association between sex and genotype for haemophilia patients due to the small number of females in this group. Analysis at this stage was not done for those attending ANCs, which comprised females all under 40 years of age. Chi squared analysis shows that there was no evidence to support the hypothesis of an association between genotype distribution and either sex or age for 5 of the 6 patient groups (p values >0.05). Log-linear regression analysis showed a highly significant interaction between group and genotype ($p = 0.0007$), suggesting that the genotype distribution differs significantly between patient groups. The genotype distribution among haemophilia patients, when compared to the other five groups as a whole, was significantly different ($p = 0.0023$).

3.4 Discussion

3.4.1 Genotyping techniques

Several different methods have been used to genotype HCV without resorting to DNA sequencing. DNA amplification with genotype-specific primers, which is usually based on the core region, is one such method (Okamoto et al. 1992; Wu et al. 1997; Ohno et al. 1997). The main advantage that this genotyping method offers is the ability to identify a large number of different subtypes. This is due to the relatively high sequence diversity in this region compared with the 5'NCR, which is the basis for many other genotyping techniques. The disadvantages of genotyping HCV with specific primers are associated

with primer annealing. Mis-priming can occur leading to incorrect genotype assignment or, more commonly, false identification of mixed infections. (Wu et al. 1997; Ohno et al. 1997).

Another commonly employed technique is hybridisation of biotinylated PCR products to membrane-bound type-specific probes (line probe assay) (Stuyver et al. 1996). This assay is based on the highly conserved sequence of the 5'NCR. Polymorphisms in the probe-binding region that might produce false genotype assignment, or no genotype at all, are unlikely. The disadvantage of such low sequence diversity within genotypes is that this method is limited in the number of different subtypes that it can identify. The line probe assay is too expensive for routine use in many laboratories.

A third popular method for HCV genotyping is RFLP analysis of amplified DNA (Davidson et al. 1995; Pohjanpelto et al. 1996). RFLP assays are also based on the 5'NCR, therefore they share many of the advantages and disadvantages of the line probe assay. However one advantage is that polymorphisms create or destroy restriction sites, often giving rise to 'unusual' band patterns rather than incorrectly typing or failing to type samples. RFLP is also more cost effective than the other two methods discussed here. The PCR-RFLP method used for this study is based on that described by Pohjanpelto et al. (1996). While this technique can distinguish between genotypes and subtypes (1a, 1b, 2a/2c, 2b, 3a, 3b, 4, 5 and 6), other subtypes could either be untypable or incorrectly assigned to one of these subtypes, and genotypes 7, 8 and 9 would be classified as genotype 1 (Tokita et al. 1998).

3.4.2 Prevalence of 3a and mixed infections in other studies

The most prevalent HCV genotypes in England and Wales were found to be 3a (37%), 1a (31%) and 1b (15%). These genotypes are found in other countries, but the finding of 3a as the most common subtype suggests that the overall distribution is more similar to northern European countries (Table 3.1). The small number of infections by genotypes that are not common to Europe reflect infection carriers who have recently migrated to Europe.

Country	Size of study	Risk group of subjects	Genotypes, ranked in descending order of frequency (%)					
			3a	1a	1b	2b	2a/c	Other
England and Wales ^[a]	567	Various	3a (37)	1a (32)	1b (15)	2b (5)	2a/c (3)	Other (8)
Finland ^[b]	264	Various	3a (41)	1b (24)	2b (20)	1a (14)	Other (1)	
Germany ^[c]	379	Chronic hepatitis patients	1b (56)	3a (26)	1a (14)	2 (4)		
Spain ^[d]	414	Chronic hepatitis patients	1b (82)	1a (8)	3a (5)	2a/c (3)	4 (1)	
Southern Italy ^[e]	148	Various	1b (51)	2c (45)	3a (3)	4 (1)	2b (1)	
USA ^[f]	179	Various	1a (58)	1b (21)	2b (13)	3a (5)	2a (2)	4 (1)
Thailand ^[g]	235	Blood donors	3a (39)	1b (20)	6 (18)	1a (9)	3b (4)	Other (10)
Taiwan ^[h]	562	Blood donors	1b (60)	2a (16)	2b (12)	3a (2)	Mixed (7)	Other (3)

Table 3.1: Comparison of HCV genotype distributions in eight countries. Only published studies with >100 subjects are included.^a This study; ^b(Pohjanpelto et al. 1996); ^c(Berg et al. 1997); ^d(Zein et al. 1996; Guadagnino et al. 1997; Lopez-Labrador et al. 1997); ^e(Kanistanon et al. 1997); ^f(Wu et al. 1997).

Only 7 (1%) samples in my study contained more than one genotype. Other groups have reported a higher incidence of infection with multiple HCV genotypes (See Table 3.2). It is seen that, generally, the highest prevalence of mixed infection becomes most apparent when genotyping is performed using type specific primers. This method has been reported to lack specificity, as type-specific primers tend to amplify other types relatively efficiently. This could lead to an over-estimation of the prevalence of mixed infection (Ohno et al. 1997; Spada et al, 1998).

Risk group	Study size	Frequency of mixed genotype or subtype (%)	Genotyping method	Reference
IDU	102	5.0	RFLP	(Stark et al. 1995)
	88	0	RFLP	(Pohjanpelto et al. 1996)
	88	4.5	Probe hybridisation	(Vitale et al. 1998)
	31	16.1	Probe hybridisation	(Garcia et al. 1998)
	90	11.0	Type specific PCR	(Silini et al. 1995)
Haemophilia patients	23	0	RFLP	(Smuts and Kannemeyer, 1995)
	48	6.3	RFLP	(Picchio et al. 1997)
	45	11.0	Probe hybridisation	(Tagariello et al. 1995)
	45	4.0	Probe hybridisation	(Tuveri et al. 1997)
	63	32.0	Type specific PCR	(Isobe et al. 1995)
	96	12.0	Type specific PCR	(Fujimura et al. 1996)
	83	3.6	Type specific PCR	(Fukuda et al. 1998)

Table 3.2: Prevalence of mixed genotype or subtype infections in IDUs and haemophilia patients. Genotyping was performed by one of three methods: RFLP, PCR followed by hybridisation of type specific probes or PCR with type specific primers.

The prevalence of mixed infections (1%) in the study described in this chapter appears low, particularly in potentially multiply exposed groups such as haemophilia patients and IDUs, and it is possible that the PCR-RFLP genotyping procedure used in this study is not as sensitive as type-specific PCR for the detection of minor strains.

Superinfection with both heterologous and homologous strains of HCV has been demonstrated in chimpanzees (Farci et al. 1992; Okamoto et al. 1994) and in humans (Kao et al. 1993; Kao et al. 1996). Such studies show that superinfection may be accompanied by a bout of acute hepatitis but that only one genotype is detected subsequently, which could be the superinfecting strain (Kao et al. 1993; Okamoto et al. 1994) or the original strain (Farci et al. 1992; Okamoto et al. 1994). The prevalence of mixed infections in my study may be low because the non-dominant type is circulating at undetectable levels. Indeed, during the course of infection, a change in the prevailing genotype can be seen in some patients, which could be due to a superinfection or takeover of the prevailing genotype in a persistent mixed infection (Jarvis et al. 1994; Pujol et al. 1998).

3.4.3 Genotype and mode of infection

The HCV genotype distribution among haemophilia patients differed significantly from the other five groups. This could be due to infection by non-indigenous strains contaminating clotting factor concentrates imported from North America, where types 1a and 1b prevail (Table 3.1), or may reflect genotypes circulating in the UK at a time prior to the heat treatment of blood products. Otherwise there was a striking similarity in HCV genotype distribution between groups, suggesting a common origin for many of these infections. Widespread injecting drug use in the community may account for this overall uniform

pattern of genotype distribution (Pawlotsky et al. 1995; Basaras et al. 1997). A recent study found that between 1 in 1000 and 1 in 2000 young adults are at risk of infection from shared drug injecting equipment (Durante and Heptonstall, 1995). During the period 1992 to 1996, in a sample of patients with HCV infection, 80% were found to be IDUs (Ramsay et al. 1998).

3.4.4 Implications of HCV genotype distribution in England and Wales

The finding that type 3a is as prevalent as type 1 has an important implication for the screening for markers of HCV infection in England and Wales. Current serological assays, based on HCV genotype 1, are reported to be suboptimal for detecting antibody to other genotypes (Neville et al. 1997). In the study by Neville et al reactivities of antibodies to a range of HCV genotypes with the components of currently used third-generation screening assays (utilising core, NS3 and NS5 antigens) were measured. Recombinant antigens expressed from different genotypes were also included in the assay to allow type-homologous and type-heterologous reactivity to be measured for each genotype. The results showed stronger reactivity to antigens of homologous type than to antigens of heterologous type. There was a delay of 88 days on average between infection with HCV and the detection of antibodies using third generation assays. This 'window' period of infection was longer for HCV infections with genotypes other than genotype 1. Thus, there is a potential for the equally prevalent type 3a to be missed during routine screening. Screening assays used in England and Wales should be optimal for genotype 3 as well as genotype 1.

HCV genotype is associated with virulence of liver disease and resistance to treatment by interferon. Individuals infected with HCV genotype 1b have a higher incidence of advanced liver disease (Dusheiko et al. 1994a; Lopez-Labrador et al. 1997; Bruno et al.

1997; Mihm et al. 1997). This observation may be due to the fact that genotype 1b is associated with a longer duration of infection than other genotypes, like 3a, which appears to have spread relatively recently with the increase in IDU (Pawlotsky et al. 1995; Basaras et al. 1997). However, the apparent association of genotype 1b with severity of liver disease could also be explained by the tendency of this genotype to resist rapid turnover, as discussed earlier. Genotype 1b tends to predominate eventually in mixed infections, suggesting that it may indeed possess an adaptive advantage. Patients infected with genotype 1b also have a poorer response to interferon therapy (Dusheiko and Simmonds, 1994b; Lin et al. 1996; Zein et al. 1996; Kiyosawa, 1997; Isaacson et al. 1997). Once again, this association could be related to the longer duration of infection in genotype 1b-infected individuals. In addition to this, advanced liver disease such as cirrhosis and hepatocellular carcinoma may reduce the effectiveness of interferon therapy.

While some studies have identified a region of the NS5a gene between amino acids 2209 and 2248, the ISDR, as an important determinant of the sensitivity of HCV genotype 1b to interferon treatment (Chayama et al. 1997; Kurosaki et al. 1997; Saiz et al. 1998), other studies have found no such link (Zeuzem et al. 1997; Duverlie et al. 1998). (Dusheiko et al. 1996; Bruno et al. 1997; Isaacson et al. 1997; Kiyosawa, 1997b; Mihm et al. 1997; Pawlotsky et al. 1998b). Continued surveillance is needed to monitor the spread of possibly more virulent and drug-resistant HCV genotypes. This will guide funding allocations for prolonged or combined antiviral therapy and liver transplantation.

Chapter 4

Investigating HCV Quasispecies Using Denaturing Gradient Gel Electrophoresis

4.1 Introduction

The HCV genome displays a high level of genetic diversity (Simmonds, 1995; Major and Feinstone, 1997; Simmonds, 1998). The genotypic variation of HCV between infected individuals has been discussed in Chapter 3. This and the next chapter describes an investigation into the intra-host variation of HCV.

Within infected individuals HCV circulates as a population of genetically distinct, but closely related, viral particles, or “quasispecies” (Domingo et al. 1985; Martell et al. 1992; Bukh et al. 1995; Farci et al. 1997). As discussed in Chapter 1, the term viral quasispecies describes a population in which individual genomes of the virus differ in one or more positions from the consensus sequence. When the quasispecies population is at equilibrium, one genome tends to be dominant. However, as this equilibrium is not stable, many genomes will have only a fleeting existence and the dominant genome may change as advantageous mutations occur (Domingo et al. 1995).

Some RNA viruses, including the closely related pestiviruses and flaviviruses, existing as quasispecies are known to be cleared effectively by the host’s immune response (Coffin, 1995; Smith et al. 1997; Pawlotsky, 1998a). Thus, not all RNA viruses existing as quasispecies establish persistent infections (Nowak et al. 1991; Korenaga et al. 1997; Domingo et al. 1998b).

Notwithstanding the quasispecies nature of HCV as the cause or consequence of chronic infection, genetic variability is still a major obstacle in vaccine design (Bukh et al. 1995; Farci et al. 1997) and may enable the virus to resist suppression or eradication with anti-

viral drugs such as interferon (Koizumi et al. 1995; Le Guen et al. 1997; Polyak et al. 1997; Toyoda et al. 1997; Pawlotsky, 1998a; Pawlotsky et al. 1998c) .

Quasispecies diversity may also influence the pathogenesis of HCV. Several studies have shown that increased quasispecies diversity is associated with the severity of liver disease (Honda et al. 1994; Koizumi et al. 1995; Gretch et al. 1996; Hayashi et al. 1997; Wyatt et al. 1998; Brambilla et al. 1998). The increased immune response to a broad spectrum of quasispecies could perhaps be responsible for this liver damage. However, as the severity of liver disease also increases with the length of infection, the apparent relationship between viral diversity and pathogenicity may simply be the result of the accumulation of mutations over time.

A close examination of HCV quasispecies may therefore allow HCV pathogenesis and its response to antivirals to be better understood. Currently, studies of HCV quasispecies distributions are done by amplifying selected portions of the genome by PCR, isolating individual species by cloning and then obtaining the DNA sequence of each clone (Ni et al. 1997; McAllister et al. 1998; Manzin et al. 1998). This procedure is laborious, and techniques that detect sequence variation without resorting to DNA sequencing have been employed to allow more rapid analysis of clones (Moribe et al. 1995; Lee et al. 1997; Sullivan et al. 1998).

I have developed a denaturing gradient gel electrophoresis technique (Fodde and Losekoot, 1994) to rapidly screen clones for genetic variation. The DGGE technique was then employed to examine the quasispecies distribution of HCV in fifteen individuals to discover how the intra-host diversity of HCV varies between individuals and whether it is

related to the mode of infection. Two genomic regions, 5'NCR and NS5b were investigated. The 5'NCR, being a highly conserved region of the genome (Bukh et al. 1992), was selected to provide information on mixed genotype infections that could not be detected by RFLP. The NS5b gene was selected because, being a variable region of the genome (Behrens et al. 1996; Lohmann et al. 1997; 1998) its sequence diversity is an index of the complexity and diversity of the viral quasispecies.

4.2 Materials and Methods

4.2.1 Patients

Fifteen individuals were selected at random from the 567 individuals studied in Chapter 3. Five of these were haemophilia patients (H069, H071, H075, H858 and H865), five were blood donors (BD244, BD259, BD268, BD424 and BD426) and five were injecting drug users (IDU189, IDU191, IDU230, IDU240 and IDU323).

4.2.2 Method

RNA was extracted from the sera or plasma of these 15 patients as described in Section 2.1.1. Fragments of both the 5'NCR and NS5b regions of the HCV genome were amplified by single round PCR as described in Sections 2.6.1 and 2.6.2. PCR products were cloned and inserts directly amplified from colonies as described in Sections 2.6.5, 2.6.6 and 2.6.7. Colony PCR products were visualised by agarose gel electrophoresis (Section 2.1.4). NS5b colony products were often derived from inserts of incorrect size, probably due to highly efficient cloning of small primer dimer sequences. Preliminary studies showed that a gel purification step (Section 2.6.3) prior to cloning eliminated this problem.

A polyacrylamide gel containing a gradient of denaturant (urea and formamide) from 30% to 70% was prepared as described in Section 2.6.9. Colony PCR products were electrophoresed through the gel as described in Section 2.6.11 and visualised as described in Section 2.6.12. The GC-clamp is attached to the colony PCR product (2.6.6 and 2.6.7) to prevent complete denaturation of the double stranded DNA and therefore allows detection of mutations in even the highest temperature-melting domain. DGGE analysis allows the genetic complexity of HCV infecting an individual to be evaluated. DNA sequencing of each variant identified by DGGE, as described in Section 2.6.13, completes the procedure by allowing the genetic distances between variants to be analysed. Analysis of sequence data was performed and genotypes assigned using the methods described in section 2.6.14. This DGGE technique was developed from the method described by Sheffield et al (1989). The key areas of development are described below.

4.2.3 Development of methodology

4.2.3.1 Selection of suitable gradient for analysis of the 5'NCR

Colony PCR products (with GC-clamp attached) were derived from cloned 5'NCR fragments of HCV infecting two individuals and the DNA sequences determined. These pre-sequenced colony products were analysed by DGGE using a gradient of 0% - 80% denaturant. Electrophoresis conditions were 200 V for 6 hr at 60°C, the running buffer (0.6X TAE) being circulated to maintain a constant temperature. The gel was stained as described in Section 2.6.12. Figure 4.1 shows the migratory positions of the colony PCR products using these DGGE conditions. These DGGE conditions failed to detect the mutations present in the clones, despite the fact that the DNA had migrated beyond the part of the gel where denaturing conditions reach those theoretically sufficient to denature the

double stranded DNA, as determined by the MELT87 program (Lerman and Silverstein, 1987). Figure 4.1 (C) shows a DGGE analysis of non-cloned amplicons derived from HCV 5'NCR. The minor quasispecies population is only visible as a smear rather than distinct bands.

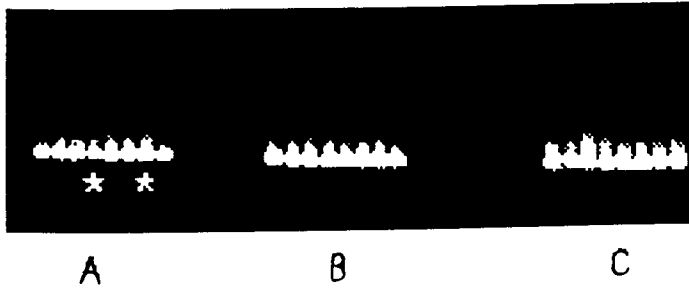


Figure 4.1: DGGE analysis of 10 µl each of: (A) eight colony PCR products derived from the 5'NCR of HCV infecting an individual, the asterisk indicates colony PCR products later shown by DNA sequencing to contain a single base substitution, (B) eight colony PCR products derived from the 5'NCR of HCV infecting another individual and (C) PCR products amplified from the 5'NCR of HCV infecting eight individuals. Electrophoresis was at 200 V and 60°C for 6 hr through a 0% - 80% gradient of denaturants.

The colony PCR products and non-cloned PCR products described above were analysed using several narrower gradients. The best result was achieved using a 30% - 70% gradient. Figure 4.2 shows the migratory positions of these PCR products through a 30% - 70% gradient. Reducing the volume of sample that was loaded onto the gel from 10 µl to 5 µl made the bands sharper. However, as the minor species present in the non-cloned PCR products (Figure 4.2 (C)) still appeared as a smear, direct DGGE analysis of PCR products could not provide useful information about the sequence diversity.

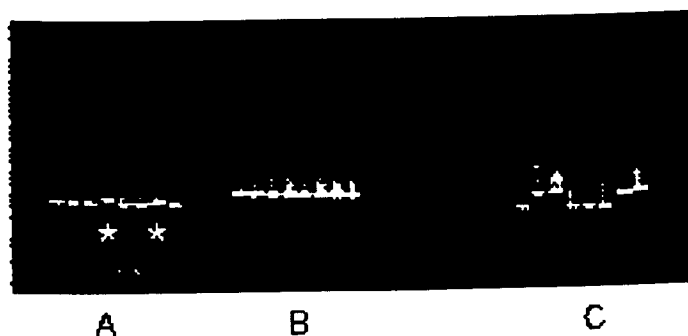


Figure 4.2: DGGE analysis of 5 μ l each of: (A) eight colony PCR products derived from the 5'NCR of HCV infecting an individual, the asterisk indicates colony products known to contain a single base substitution, (B) eight colony PCR products derived from the 5'NCR of HCV infecting another individual and (C) PCR products amplified from the 5'NCR of HCV infecting eight individuals. Electrophoresis was at 200 V and 60°C for 6 hr through a 30% - 70% gradient of denaturants.

These optimal conditions for DGGE analysis of colony PCR products derived from HCV 5'NCR were used to analyse approximately 100 more clones containing a variety of single and multiple base substitutions, pre-determined by DNA sequencing, in order to evaluate the ability of the technique to reliably detect mutations. All mutations in this test panel were differentiated by this DGGE technique (data not shown).

4.2.3.2 DGGE analysis of the hyper-variable region 1 (HVR1)

A fragment of the HVR1 was amplified as described in Section 2.6.4. Colony PCR products were generated as described in Sections 2.6.5 and 2.6.8. Figure 4.3 shows the migratory positions of 44 HCV HVR1 colony PCR products following DGGE analysis using the conditions optimised for the analysis of 5'NCR clones. This region was, however, not selected for use in subsequent studies as the degree of hypervariability was too high, and sequencing of a large number of variants would be needed to provide the necessary divergence data.

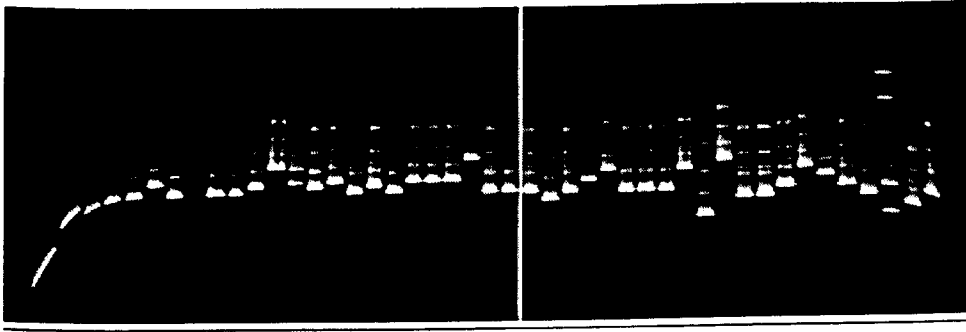


Figure 4.3: DGGE analysis of 44 colony PCR products derived from the HVR1 of HCV infecting an individual.

4.2.3.3 DGGE analysis of the NS5b region

A fragment of the NS5b region was amplified as described in Section 2.6.2. Colony PCR products were generated as described in Sections 2.6.5 and 2.6.7, but 35 cycles were performed. Figure 4.4 shows the migratory positions of 30 HCV NS5b colony PCR products following DGGE analysis using the conditions optimised for the analysis of 5'NCR clones. The smeary result could not be improved simply by loading a smaller aliquot of the sample onto the gel. However, reducing the number of cycles in the colony PCR from 35 to 20 eliminated this smearing completely. Figure 4.5 shows data from DGGE analysis of 5 μ l each of 37 HCV NS5b colony PCR products generated using only 20 cycles.

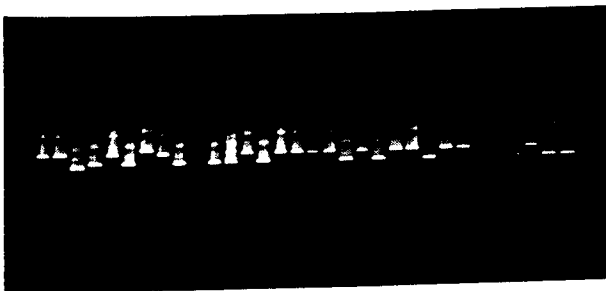


Figure 4.4: DGGE analysis of 30 colony PCR products derived from the NS5b region of HCV infecting an individual (35 cycles).

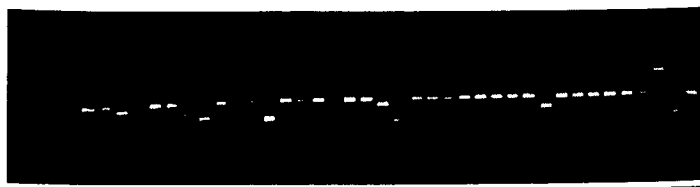


Figure 4.5: DGGE analysis of 37 colony PCR products derived from the NS5b region of HCV infecting an individual (20 cycles).

4.2.3.4 Electrophoresis conditions

Theoretically, once a double stranded DNA fragment with a GC-clamp attached has reached a point in the gradient gel where all but the clamp region has denatured it will no longer be able to migrate through the gel matrix. Consequently, I decided to increase the electrophoresis time for the DGGE procedure to 16 hr (overnight) and reduce the voltage to 100 V. I found that this did not affect the quality of the gels and made the whole procedure more streamlined.

4.2.3.5 Simultaneous analysis of both genomic regions

The final migratory positions of the 5'NCR and NS5b colony PCR products were different enough to allow 40 colony PCR products from each region to be analysed on the same gradient gel. Figure 4.6 shows DGGE analysis of 80 colony PCR products derived from the 5'NCR (lower bands) and the NS5b (upper bands) of HCV infecting a single individual using the optimised procedure described in Sections 2.6.1 – 2.6.12.

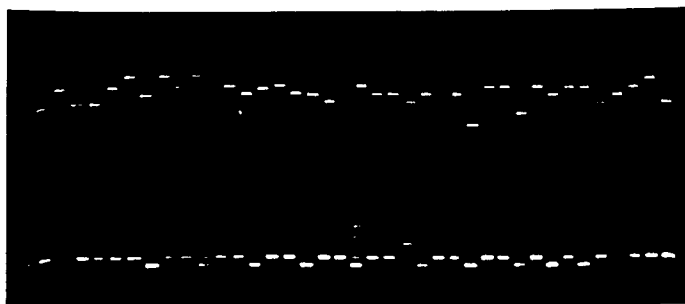


Figure 4.6: Simultaneous DGGE analysis of 40 colony PCR products derived from the 5'NCR (lower bands) and 40 colony PCR products derived from the NS5b region (upper bands) of HCV infecting an individual, using procedure described in Sections 2.6.1 – 2.6.12.

4.2.3.6 Investigation of heteroduplex bands

DGGE analysis, particularly of the 5'NCR products, revealed the presence of multiple species in some colony PCR products. Figure 4.7 illustrates this phenomenon. The asterisk indicates those lanes where multiple bands, in most cases four, are seen. The lower two bands may arise from different sequences and the upper two bands appeared to be the heteroduplexes formed between these two sequences, melting at a much lower temperature due to the presence of the mismatched bases. DNA sequencing of such colony PCR products confirmed the presence of two different sequences.

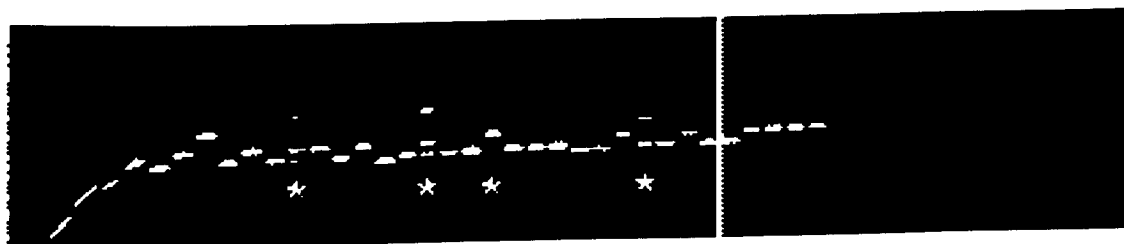


Figure 4.7: DGGE analysis of colony PCR products derived from the 5'NCR of HCV infecting an individual. The * indicates lanes where more than one species is present.

This “mixture” of sequences could have occurred at the colony PCR stage by inadvertent inoculation of the colony PCR reaction mix with more than one colony. To test this hypothesis, I selected six 5’NCR colony PCR products in which the phenomenon was observed (Figure 4.8 (A)) and streaked out the colonies from which they originated on agar plates containing ampicillin. Six of the resulting colonies were picked and amplified as described in Section 2.6.6. These colony PCR products were then analysed by DGGE. Figure 4.8(B) shows the result for two such clones, labelled 37 and 39. All six colonies gave a similar result – multiple bands were still present in each of the colony PCR products derived from the streaked out colonies. This confirmed that two different species can be amplified from a single colony.

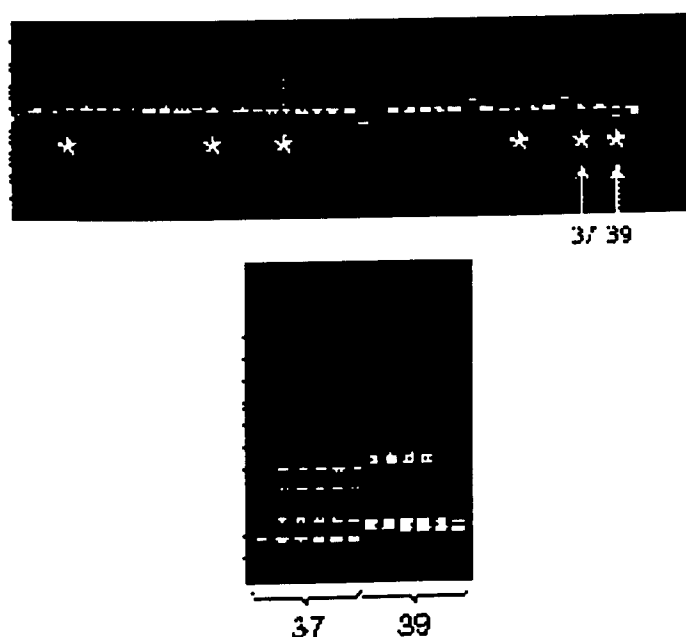


Figure 4.8: (A) DGGE analysis of colony PCR products derived from the 5’NCR of HCV infecting an individual. The * indicates lanes where more than one species is present. The colonies from which these PCR products were derived were “streaked” and PCR product amplified from six of the resulting colonies and analysed by DGGE. (B) shows the results obtained with clones 37 and 39.

I posited two possible mechanisms by which multiple bands could occur. The first of these is “superinfection” of the competent *E.coli* with two or more plasmids during the transformation stage of the cloning procedure. If the plasmids contained inserts of differing sequence then colony PCR would amplify both versions, and heteroduplexes between the two species would form during the later stages of PCR when template strands denature and reanneal to other abundant template strands rather than to the depleted primer. The second mechanism involves dimer formation when PCR product is ligated into the cloning vector. Two separate sequences would form one double-size insert containing two forward and two reverse priming sites. Colony PCR would amplify the two species separately as described above for the “superinfection” model. Amplifying two smaller sequences in this way would be more efficient than amplifying the double-size insert.

Both of these models fit the observations well, and explain why this phenomenon is seen almost exclusively in 5’NCR colony PCR products. The fragment amplified from the 5’NCR is smaller (174 base pairs) than the fragment amplified from the NS5b (401 base pairs). The cloning strategy that I used preferentially clones smaller fragments with greater efficiency than larger ones. In the context of the “superinfection” model NS5b fragments would be cloned into the vector less efficiently than 5’NCR fragments therefore producing less plasmids and reducing the probability of multiple transformations of competent cells. For NS5b fragments to form dimers, fragments of at least 802 base-pairs would have to be ligated into the cloning vector; this would be less efficient than cloning 5’NCR fragment dimers of 350 base-pairs.

Consequently, for this study, two homoduplex bands appearing in one lane, along with heteroduplex bands, were considered as having arisen from two separate sequences.

4.3 Results

4.3.1 DGGE and nucleotide sequences

The quasispecies composition of HCV infecting the 15 study individuals described in Section 4.2.1 was studied using the DGGE technique described in Section 2.6. A total of 80 colony PCR products, 40 derived from the 5'NCR and 40 from the NS5b, were analysed for each individual. The resulting gel pictures are displayed in Figures 4.9 – 4.23. Different band positions indicate differences in nucleotide sequence, each variant or “clonotype” having migrated to a unique position in the gel. Nucleotide sequences were obtained for each clonotype identified by DGGE, and are displayed, along with the frequency with which they occurred, in Figures 4.9 – 4.23. Amino acid sequences were obtained from the NS5b clonotypes and are displayed in Figures 4.24 – 4.38.

Initial analysis of the gel pictures and nucleotide sequences yielded three main observations: For the majority of individuals there was a greater number of NS5b clonotypes than 5'NCR clonotypes, with more sequence diversity between them. The quasispecies distribution differed between individuals, but there did not appear to be a trend associated with the patient group (haemophilia patients, blood donors and IDUs). Lastly, in the majority of cases, one clonotype prevailed, but in a few individuals co-dominant clonotypes could be seen (Figures 4.12, 4.15, 4.18, 4.20 and 4.22).

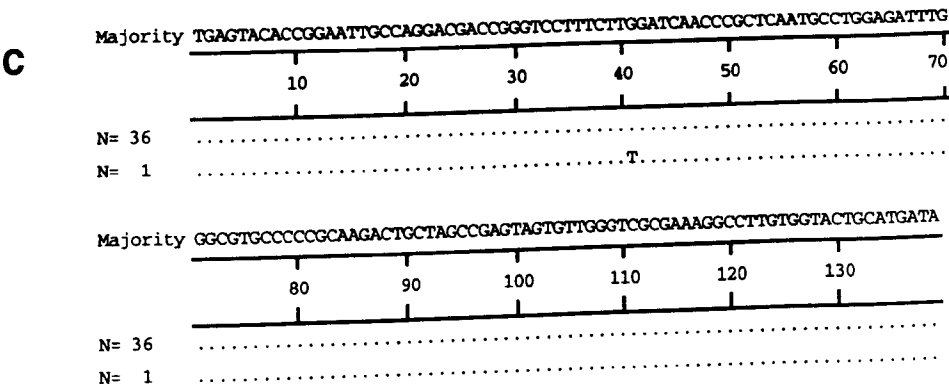
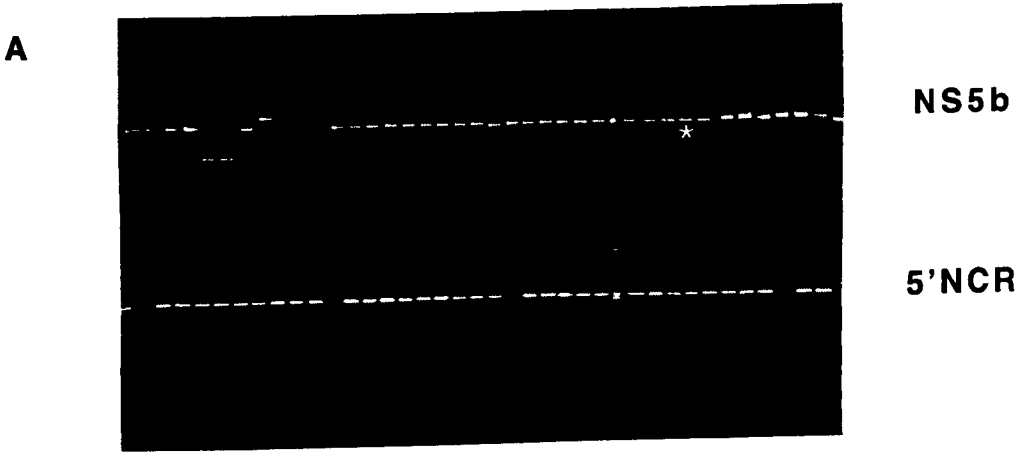
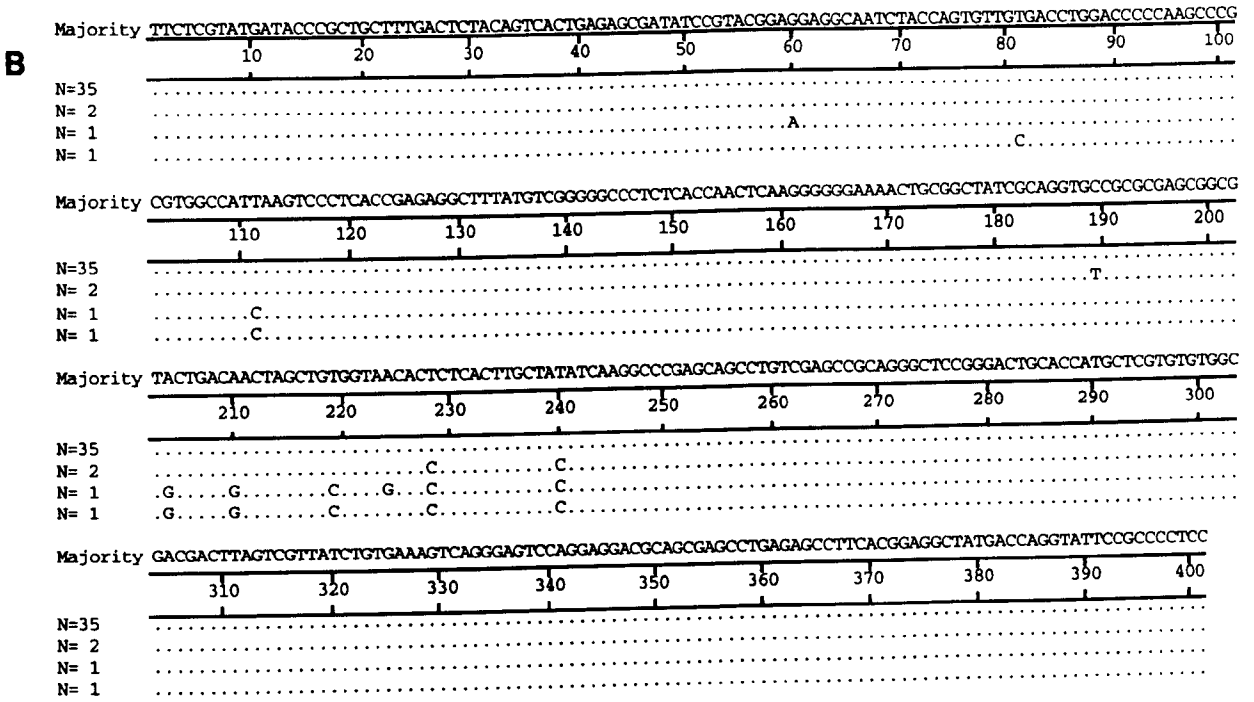


Figure 4.9: (A) DGGE analysis of clonotypes infecting patient H069. The gel shows the quasiespecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient H069. N denotes the frequency of each species. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient H069. N denotes the frequency of each species.

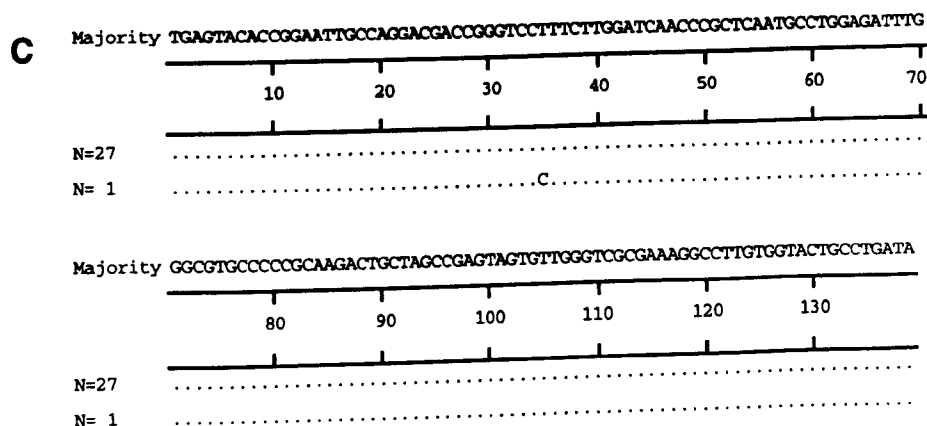
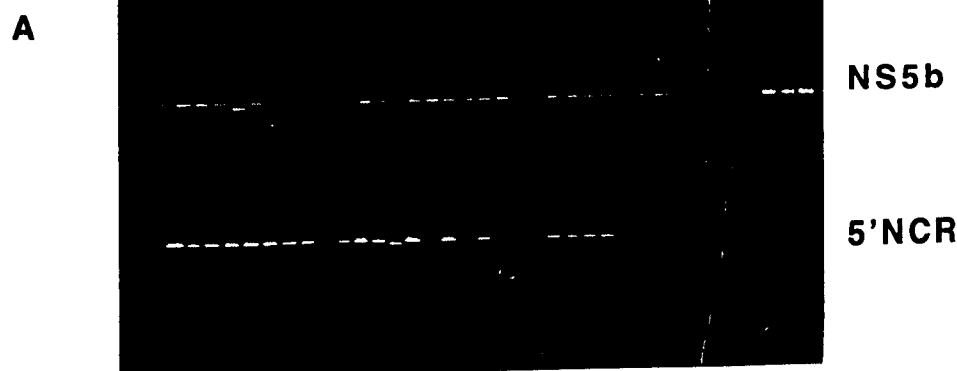
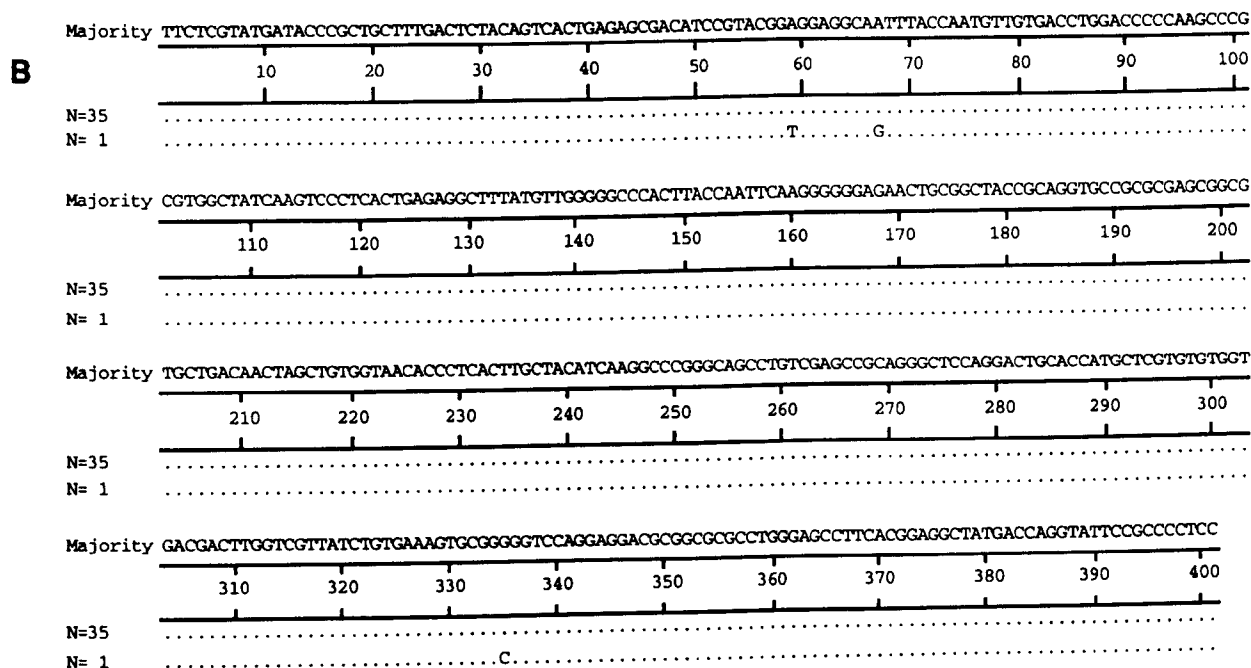


Figure 4.10: (A) DGGE analysis of clonotypes infecting patient H071. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). (B) Nucleotide sequences of NS5b clonotypes infecting patient H071. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient H071.

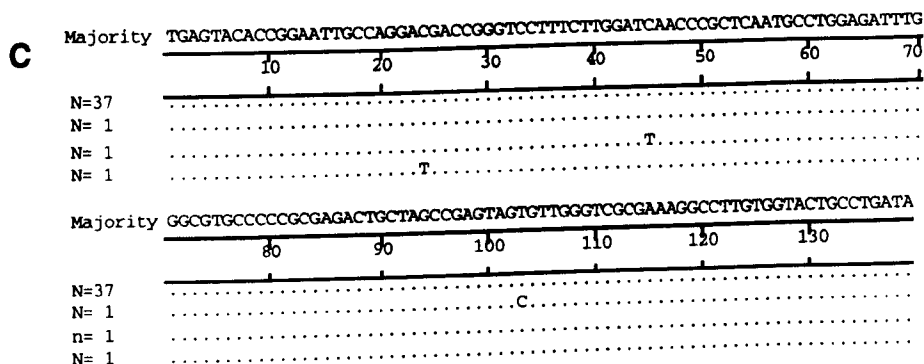
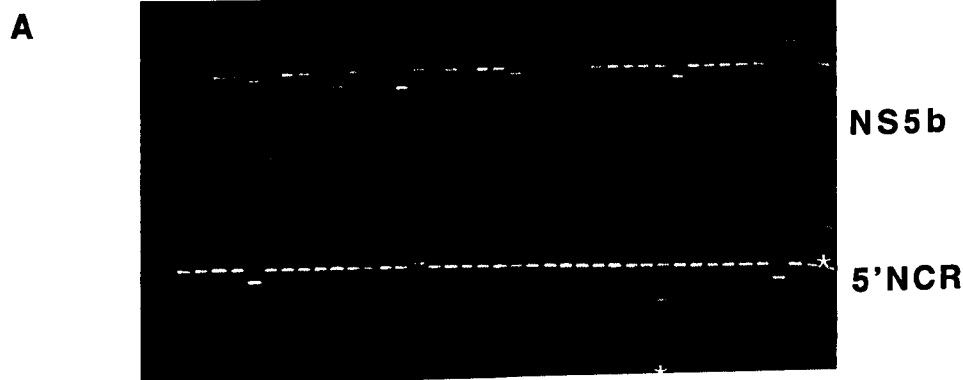
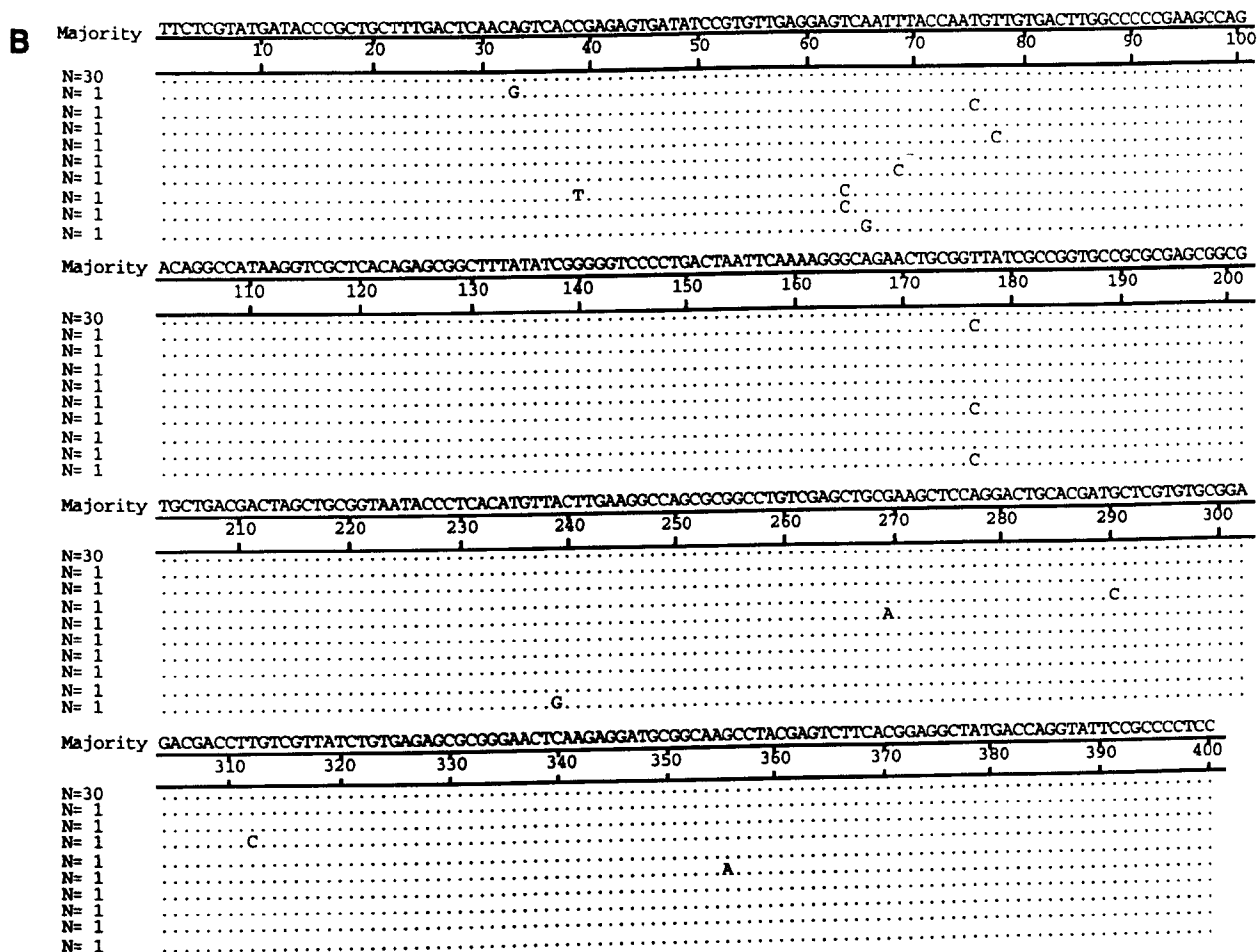


Figure 4.11: (A) DGGE analysis of clonotypes infecting patient H075. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient H075. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient H075.

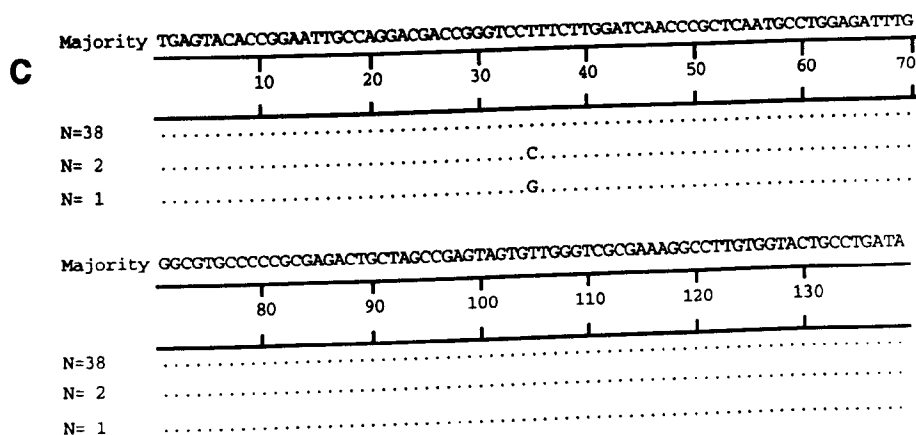
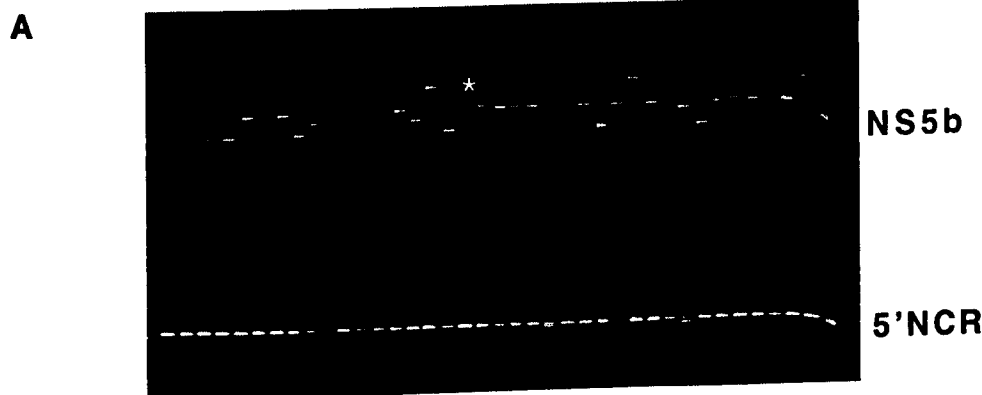
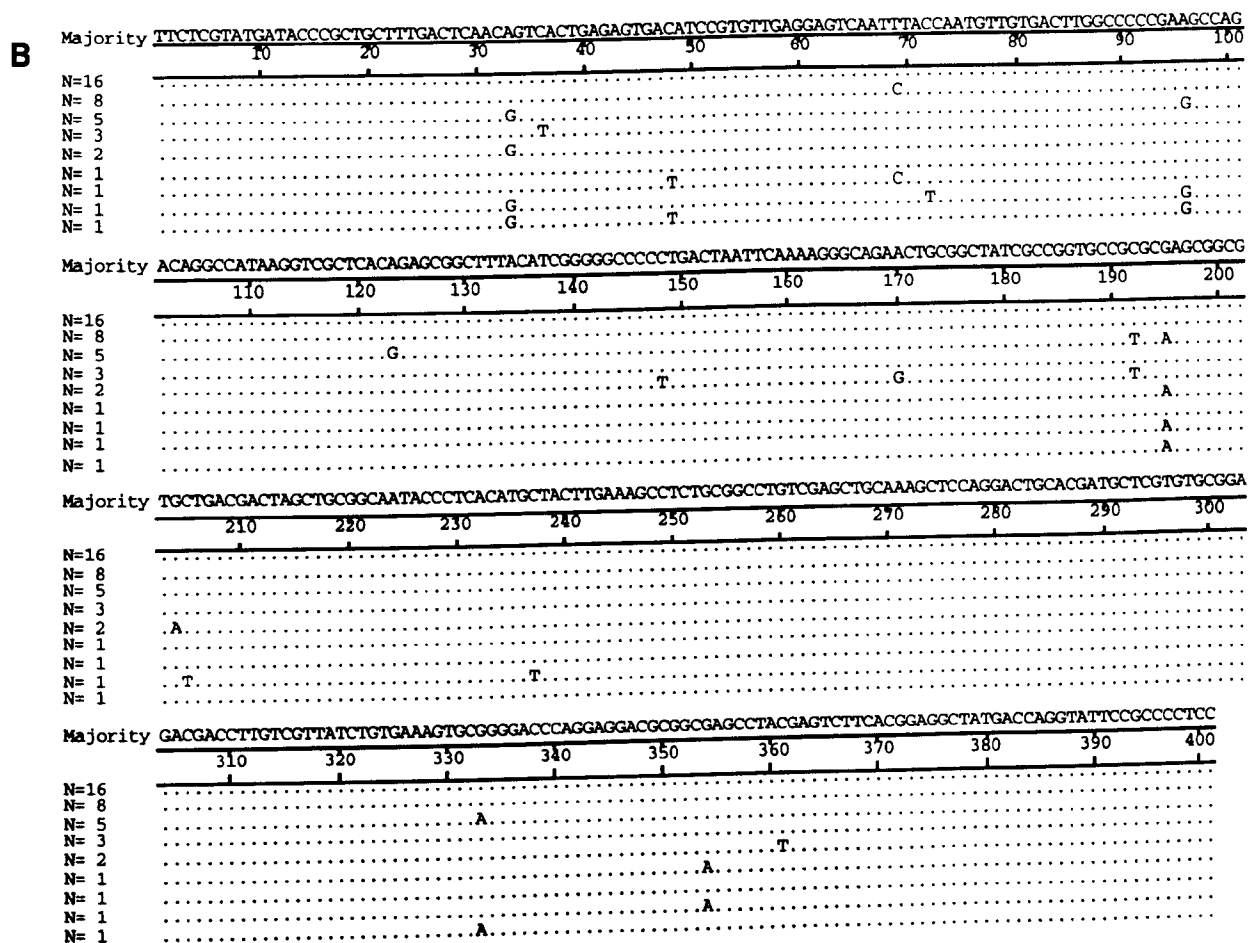
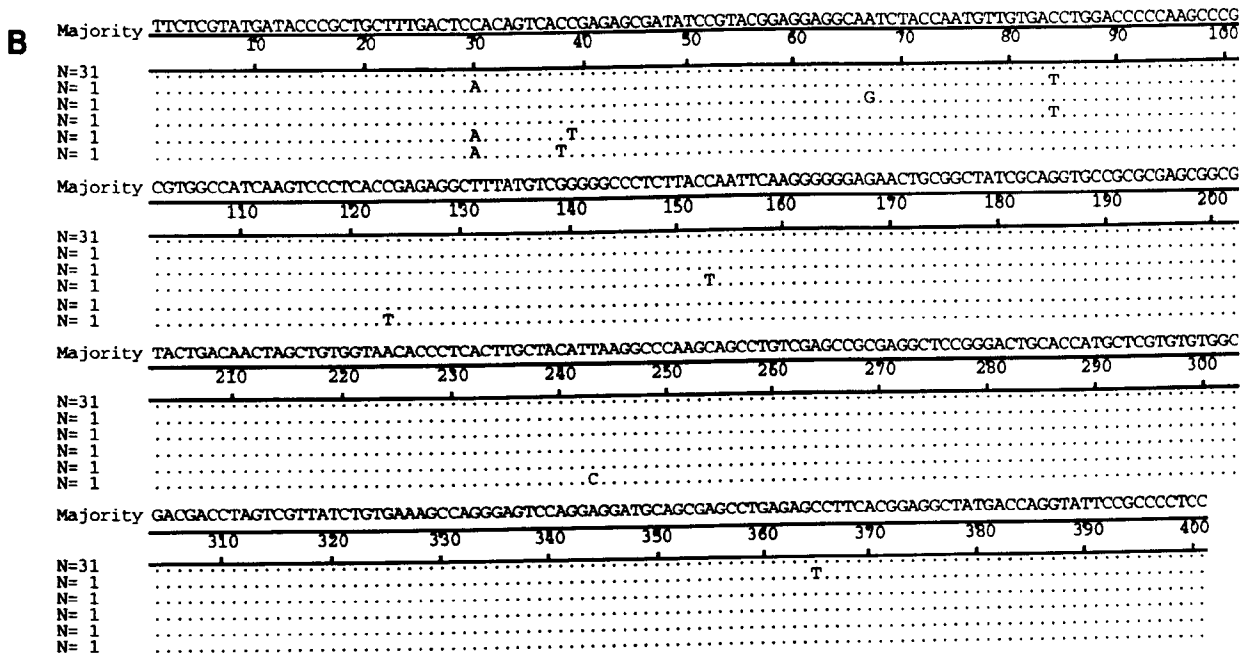
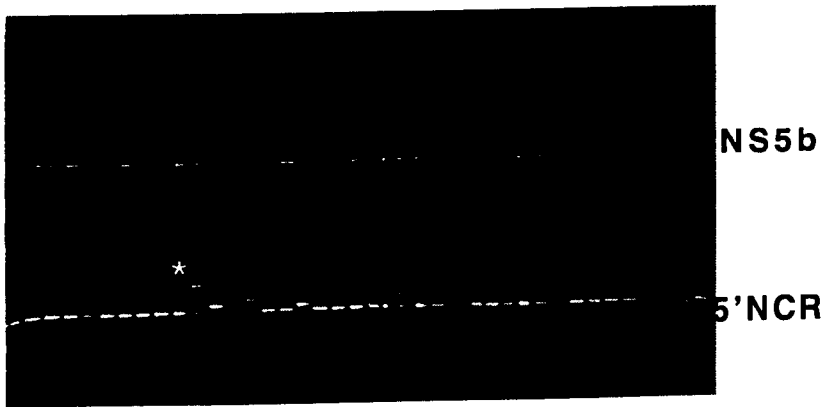


Figure 4.12: (A) DGGE analysis of clonotypes infecting patient H858. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient H858. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient H858.



A



C

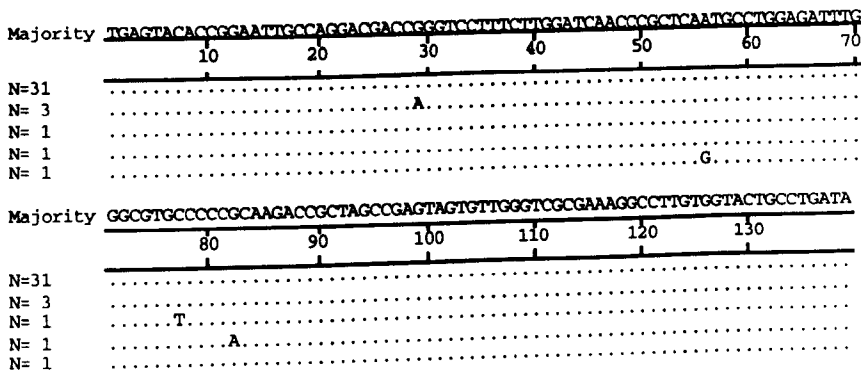


Figure 4.13: (A) DGGE analysis of clonotypes infecting patient H865. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient H865. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient H865.

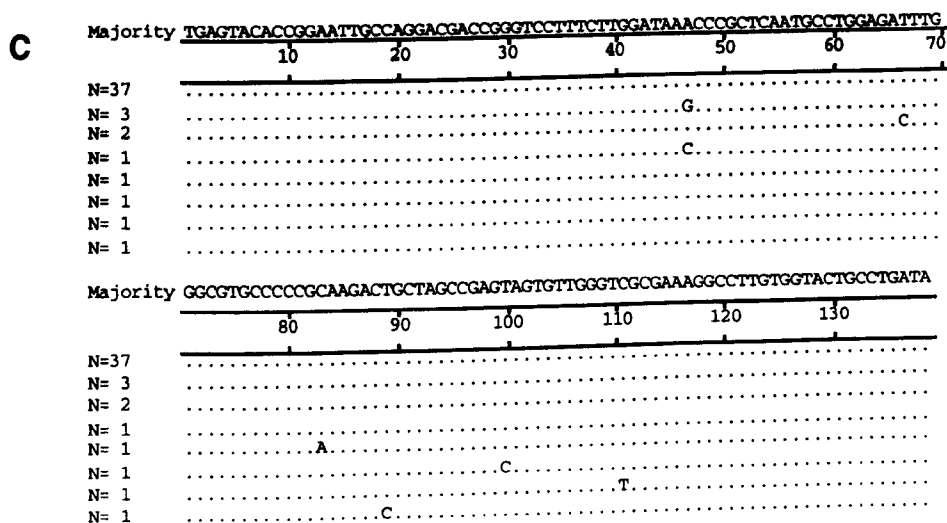
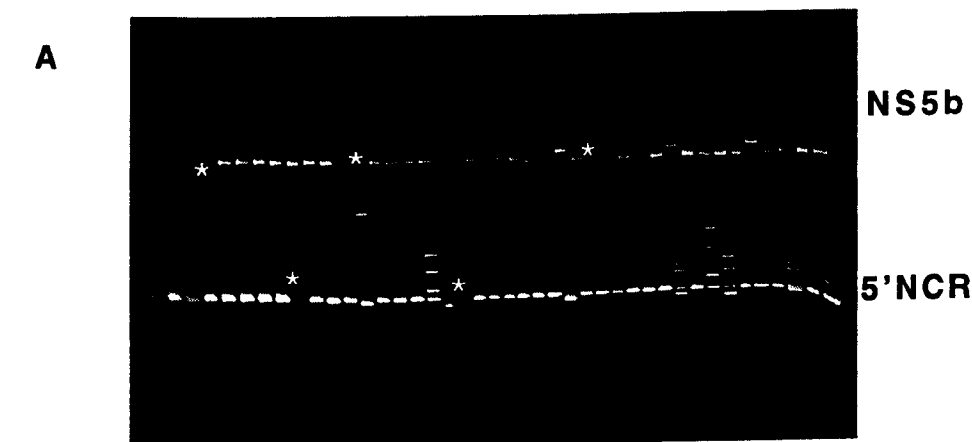
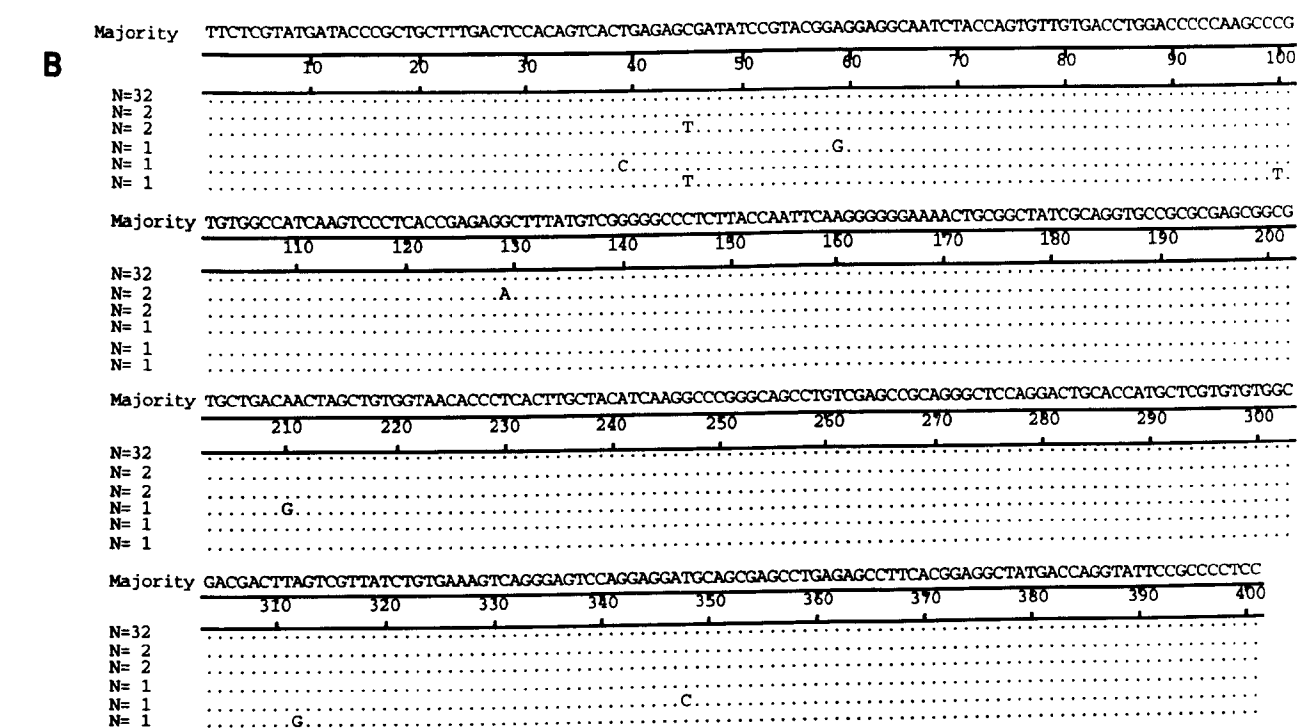
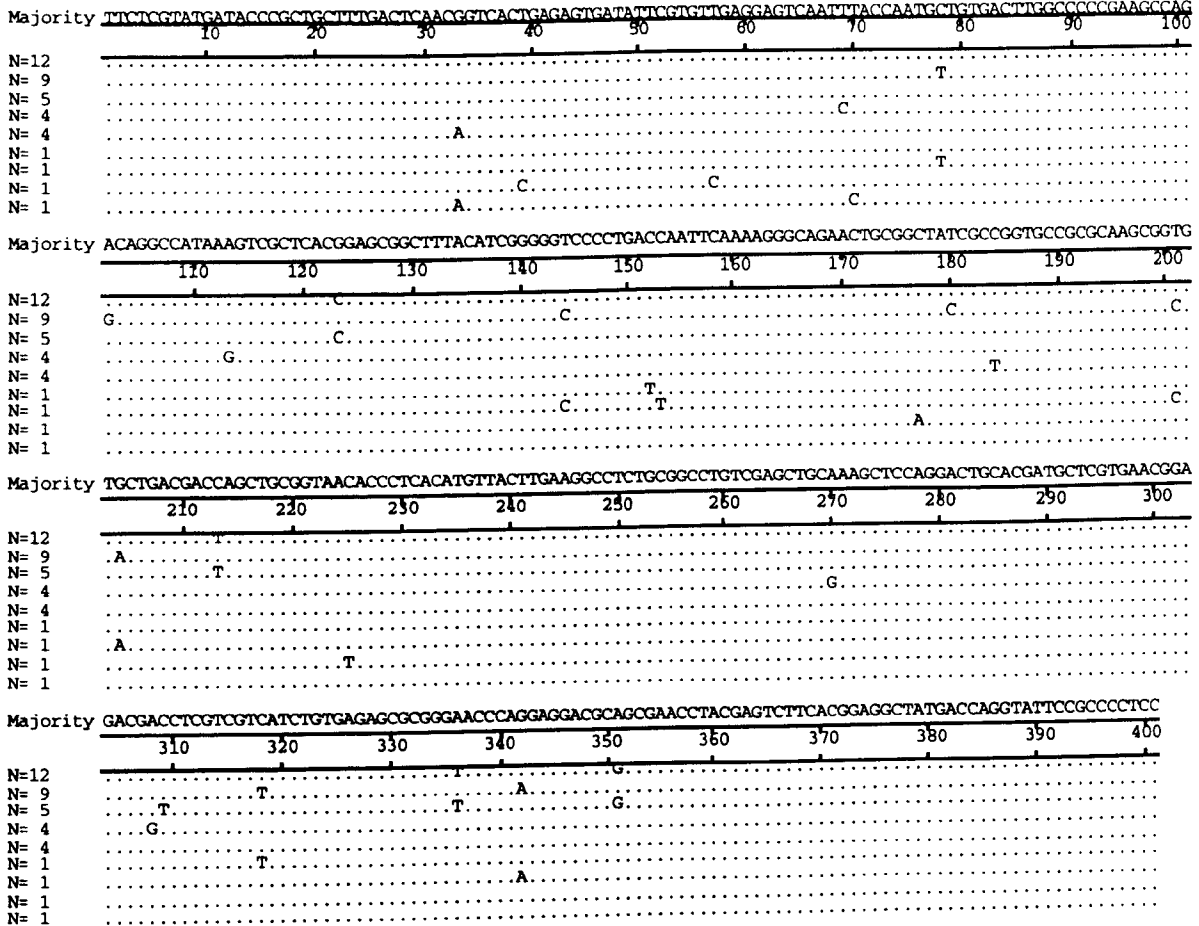
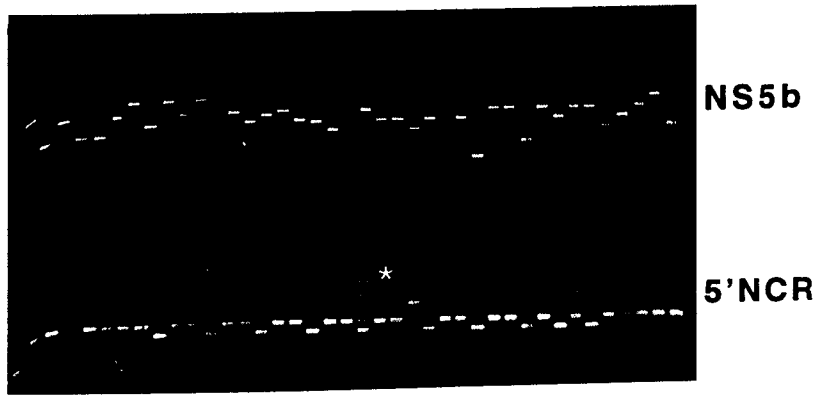


Figure 4.14: (A) DGGE analysis of clonotypes infecting patient BD244. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient BD244. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient BD244.

B



A



C

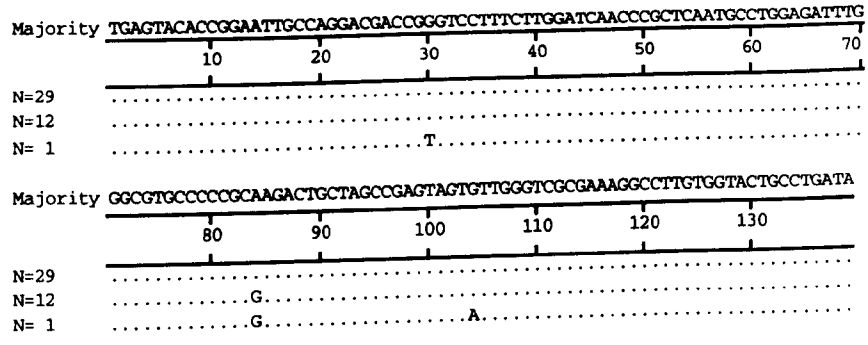


Figure 4.15: (A) DGGE analysis of clonotypes infecting patient BD259. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient BD259. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient BD259.

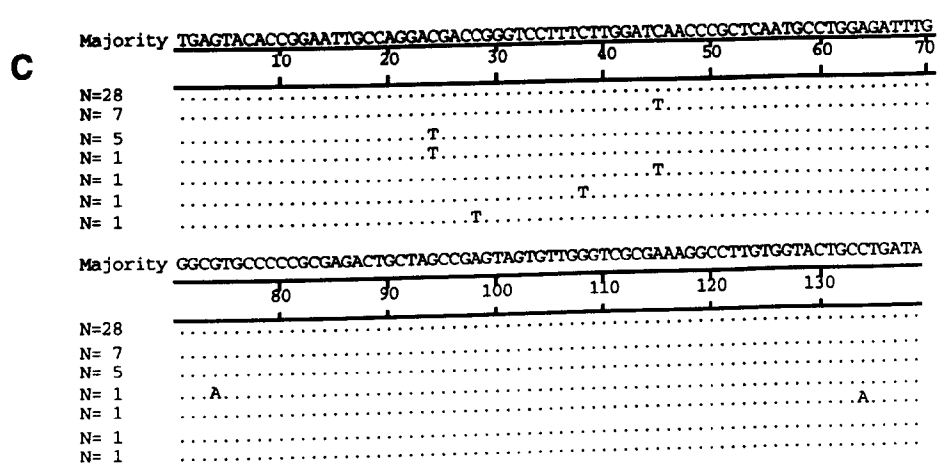
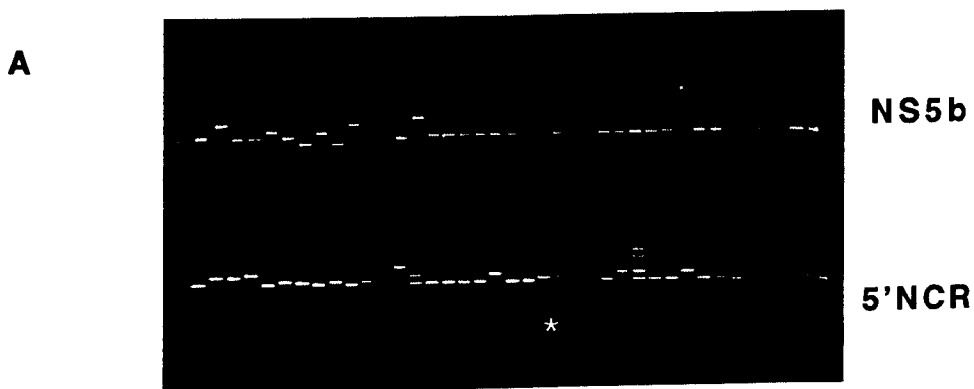
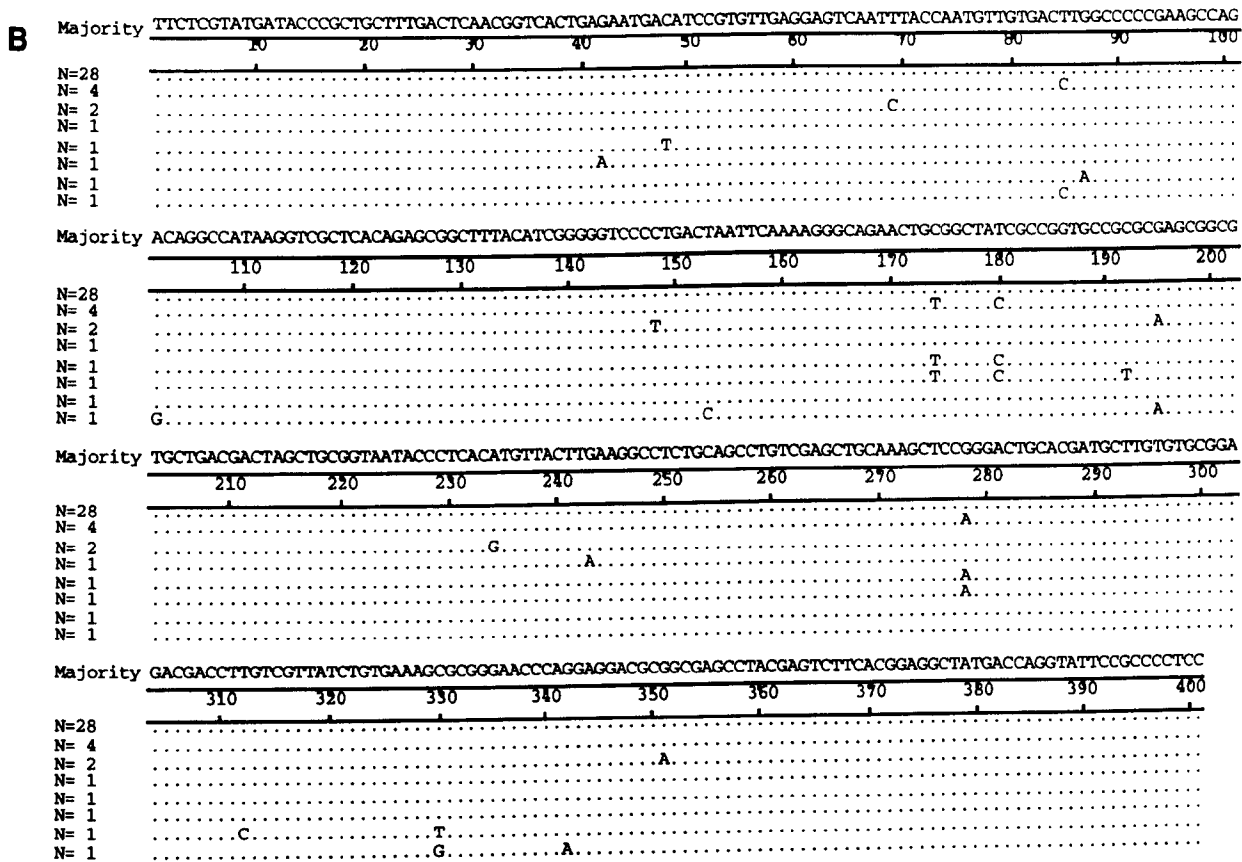


Figure 4.16: (A) DGGE analysis of clonotypes infecting patient BD268. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient BD268. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient BD268.

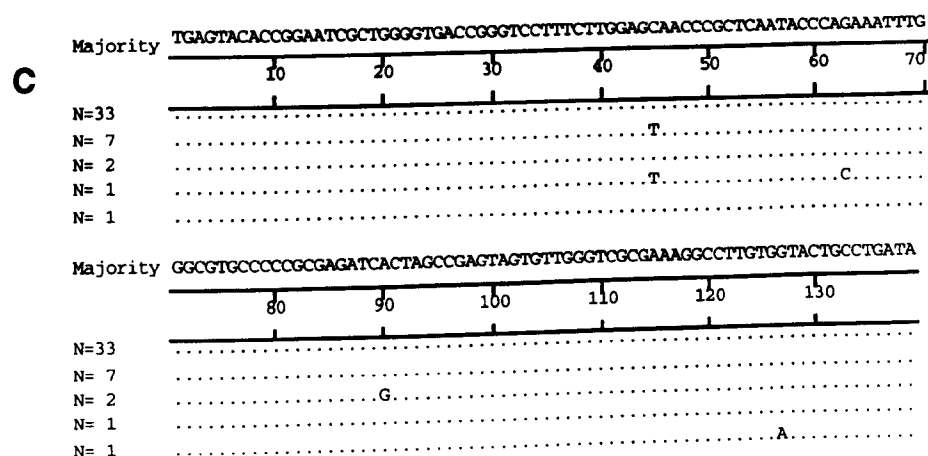
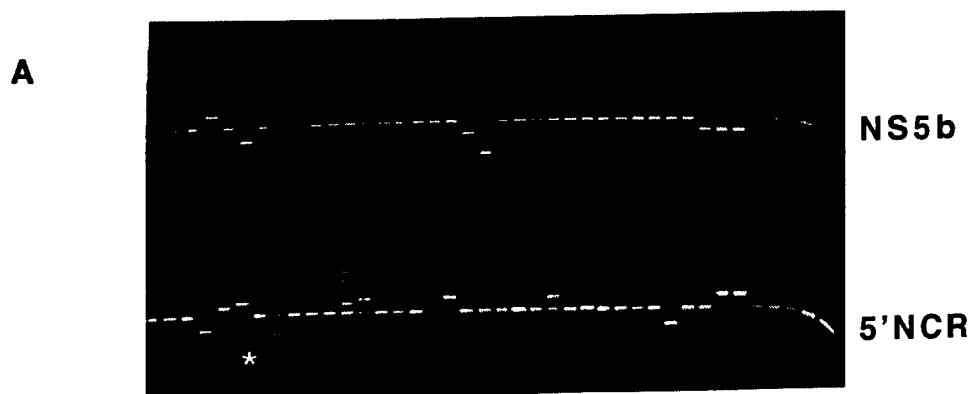
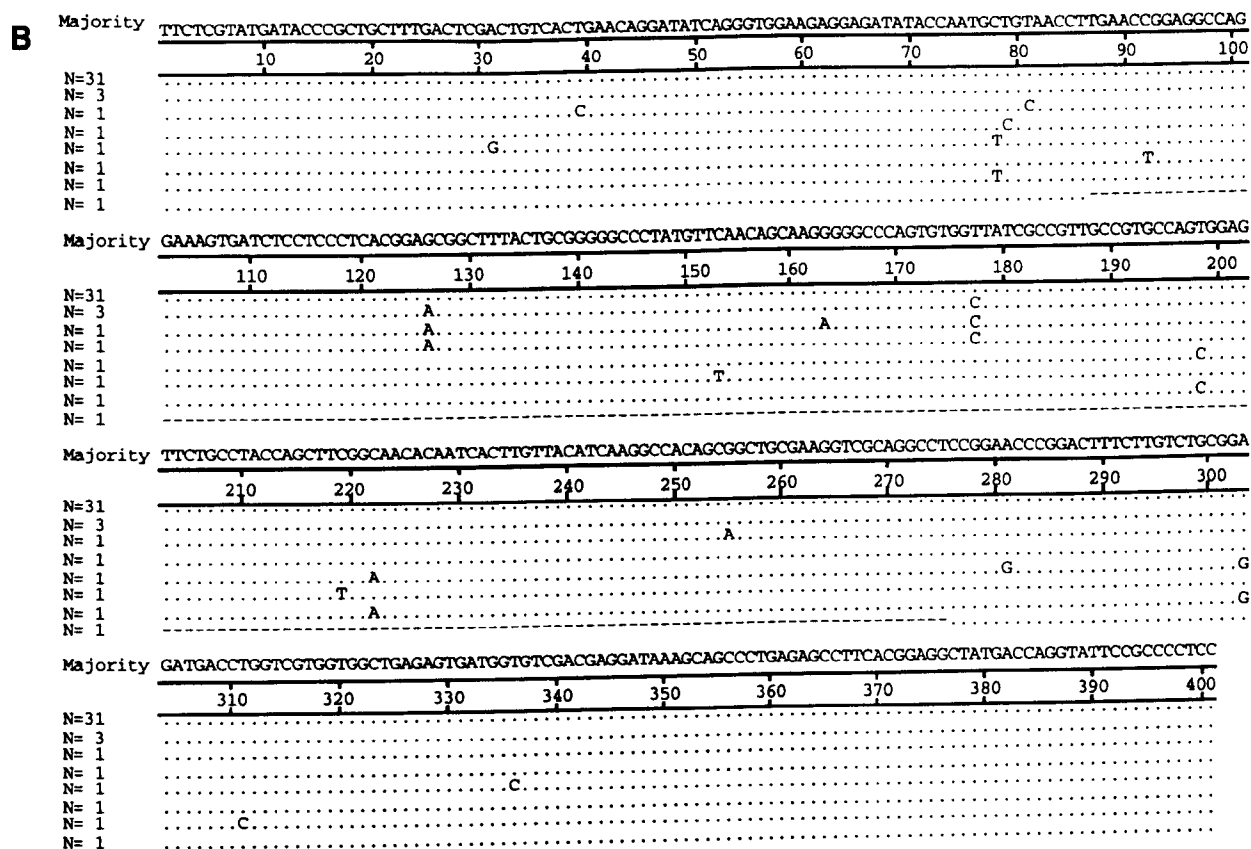


Figure 4.17: (A) DGGE analysis of clonotypes infecting patient BD424. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient BD424. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient BD424.

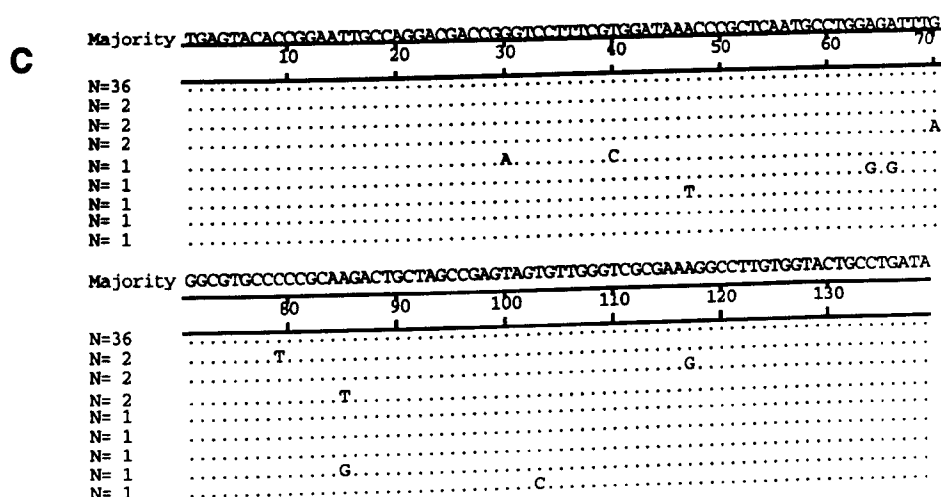
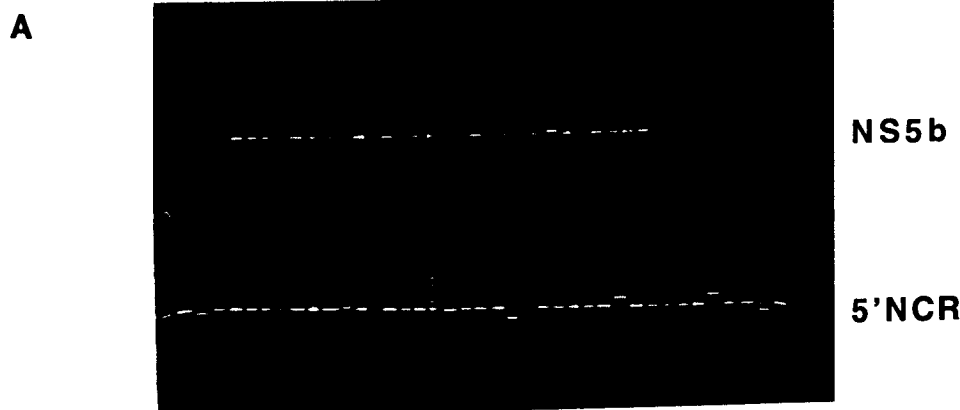
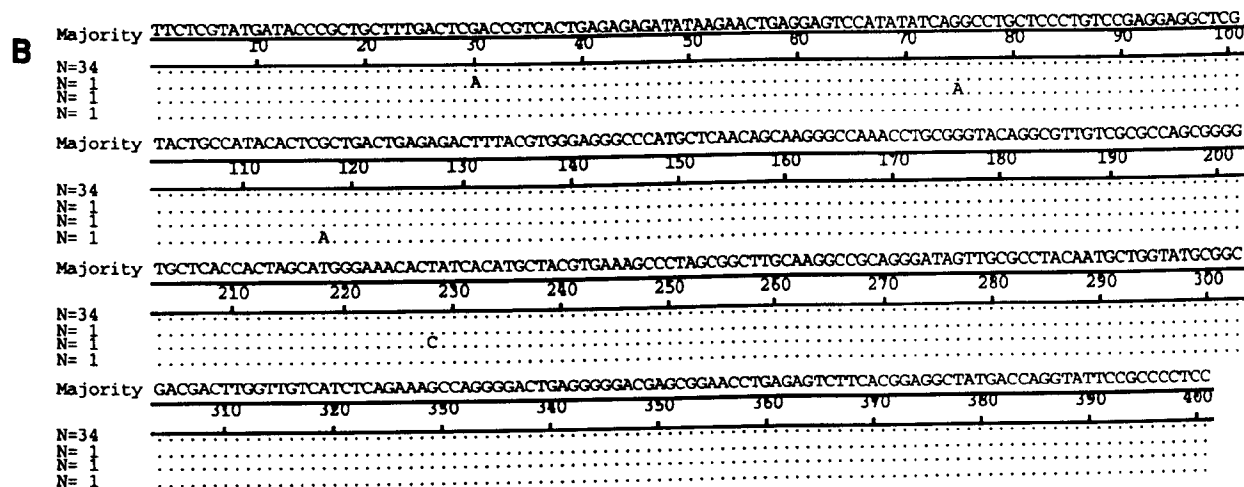


Figure 4.19: (A) DGGE analysis of clonotypes infecting patient IDU189. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). (B) Nucleotide sequences of NS5b clonotypes infecting patient IDU189. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient IDU189.

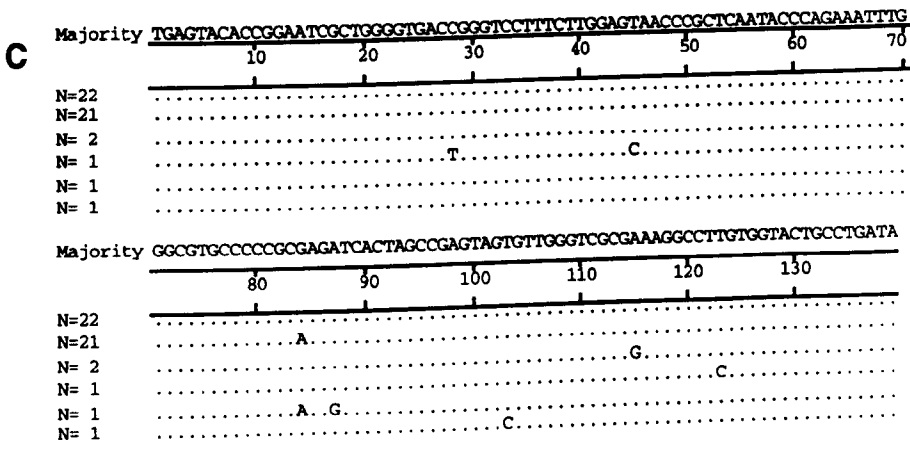
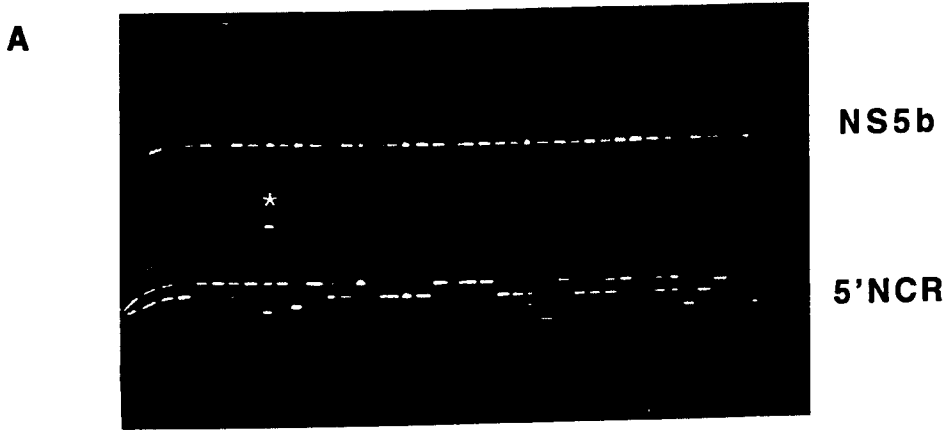
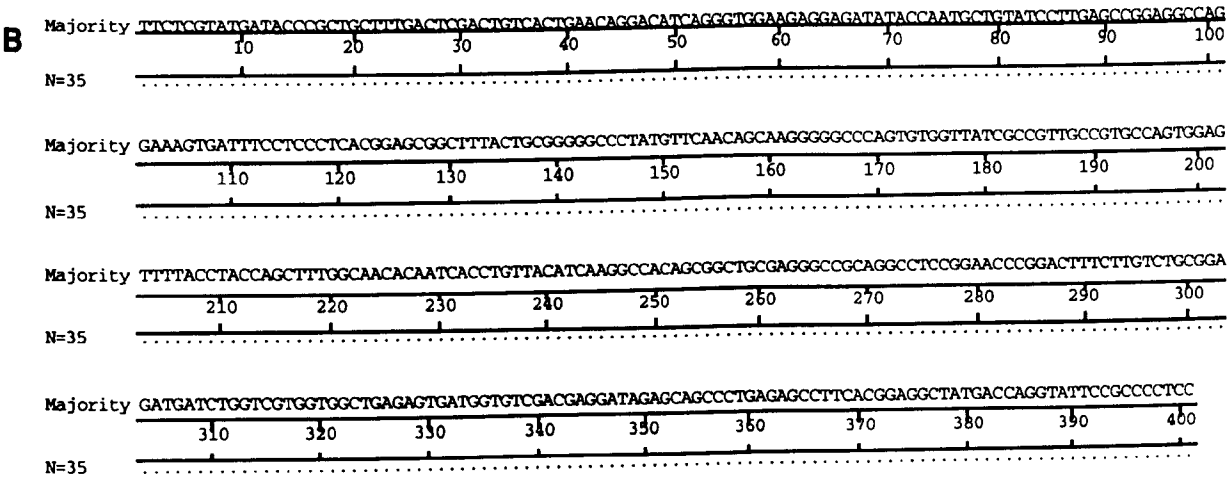


Figure 4.20: (A) DGGE analysis of clonotypes infecting patient IDU191. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient IDU191. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient IDU191.

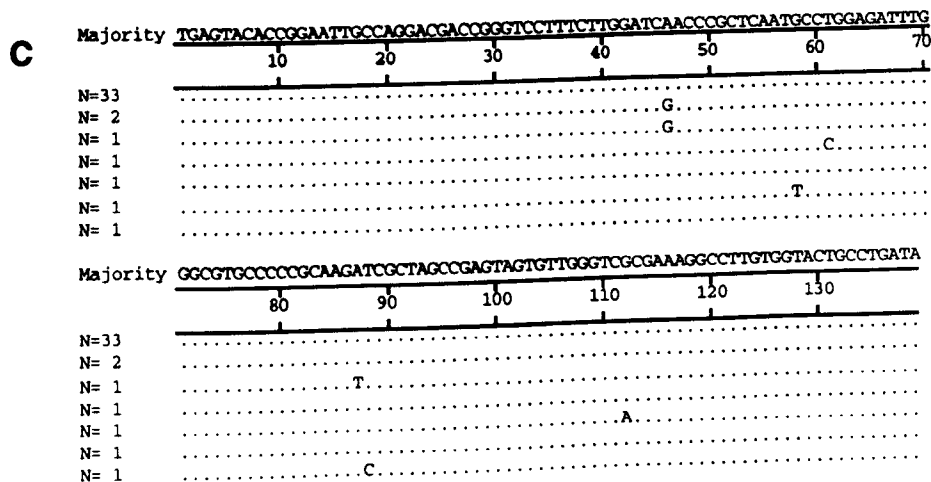
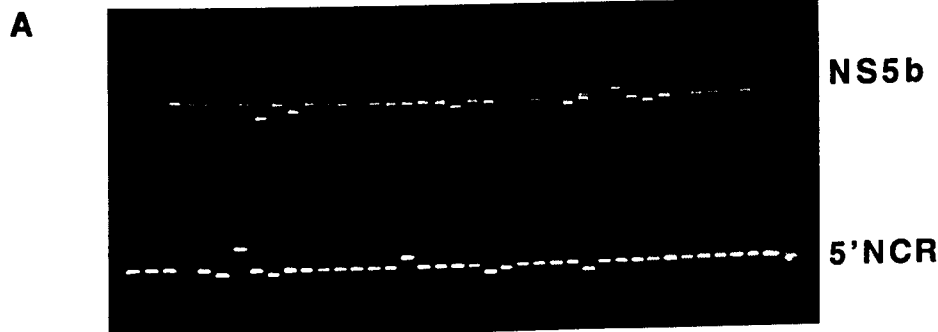
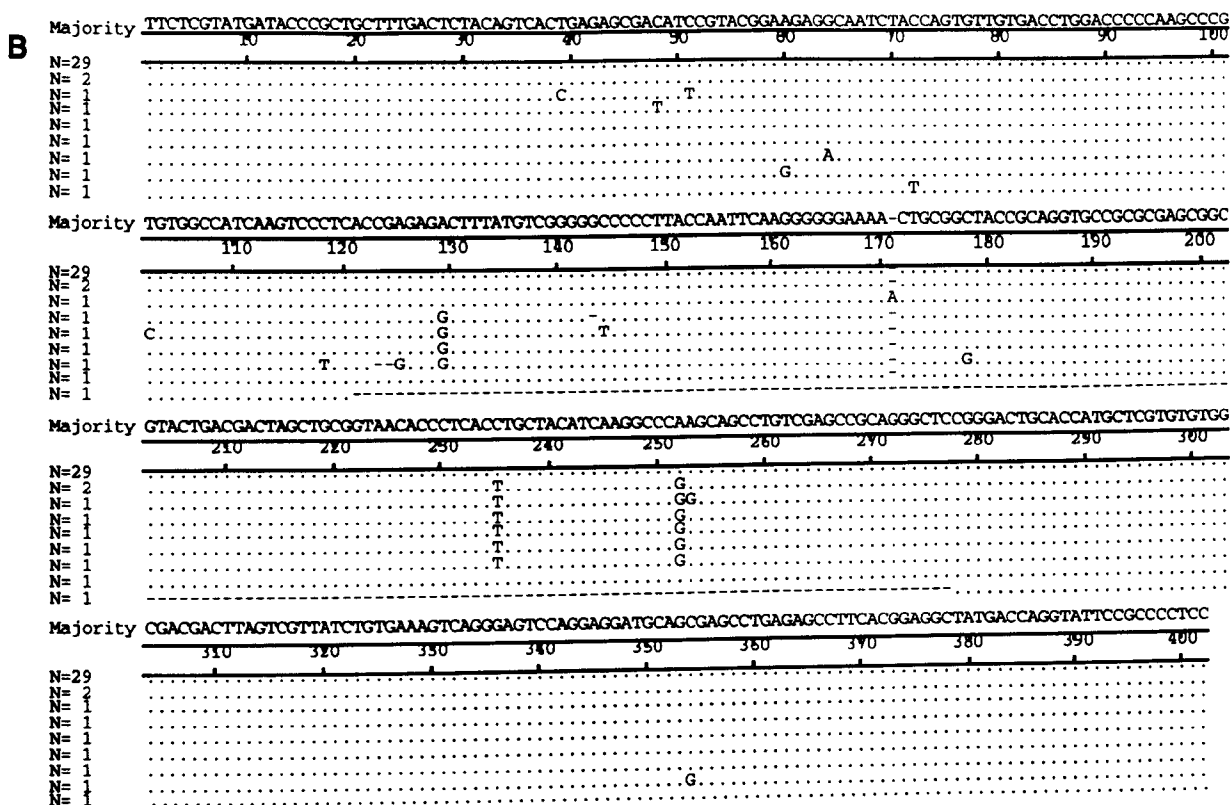


Figure 4.21: (A) DGGE analysis of clonotypes infecting patient IDU230. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). (B) Nucleotide sequences of NS5b clonotypes infecting patient IDU230. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient IDU230.

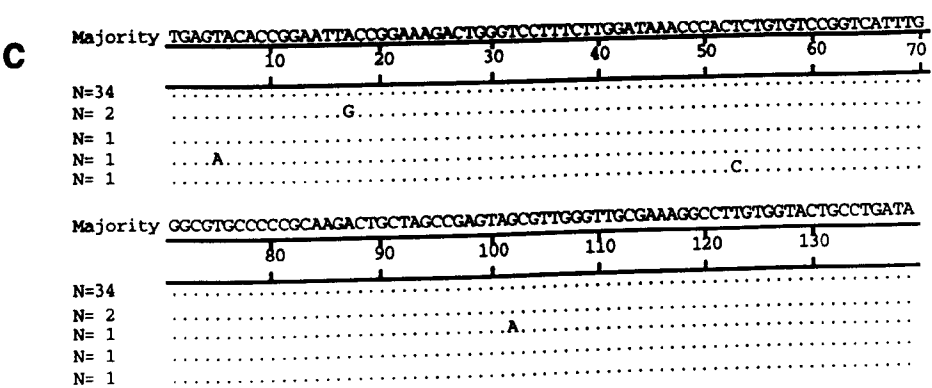
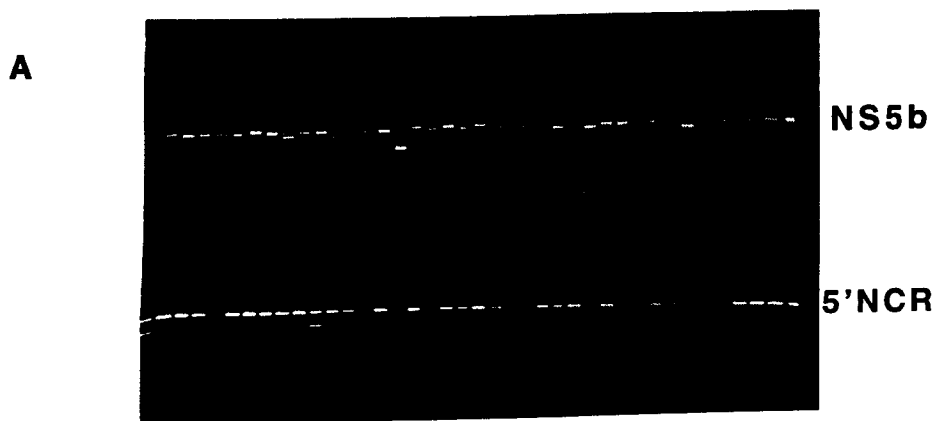
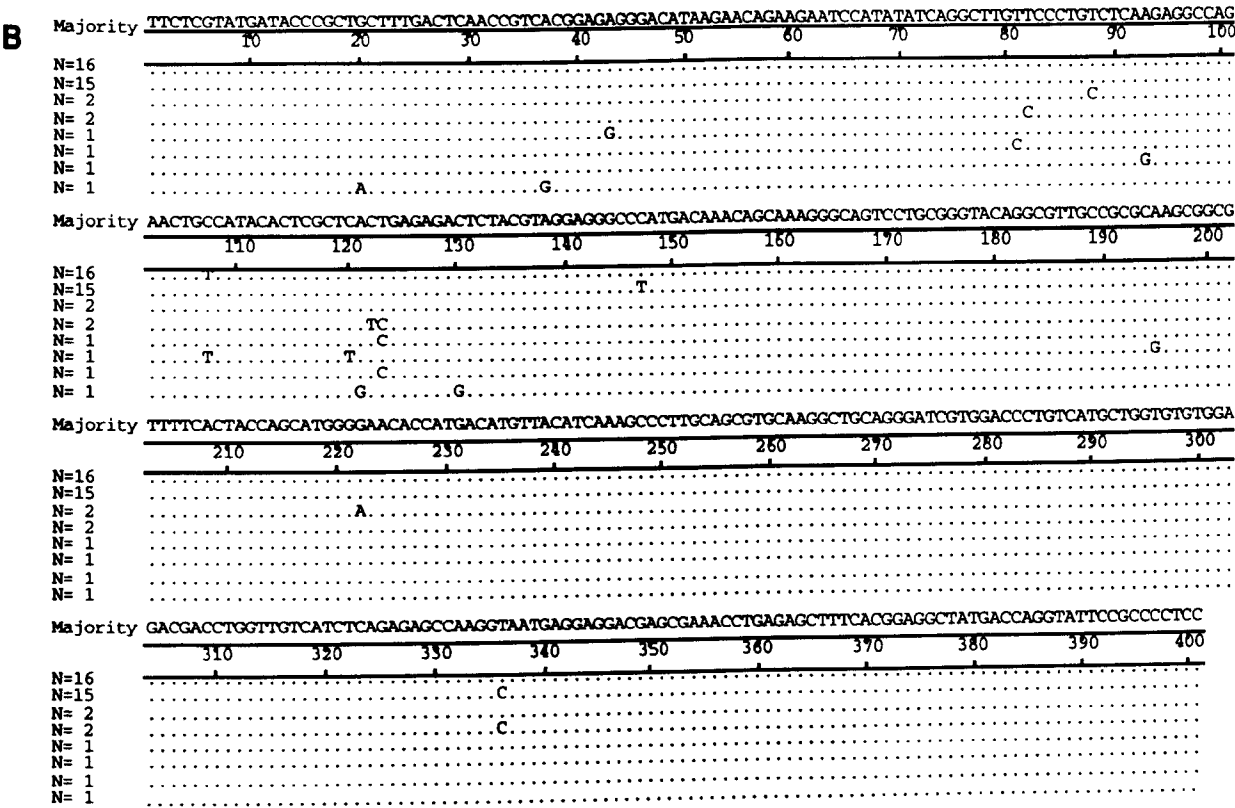


Figure 4.22: (A) DGGE analysis of clonotypes infecting patient IDU240. The gel shows the quasiespecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). (B) Nucleotide sequences of NS5b clonotypes infecting patient IDU240. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient IDU240.

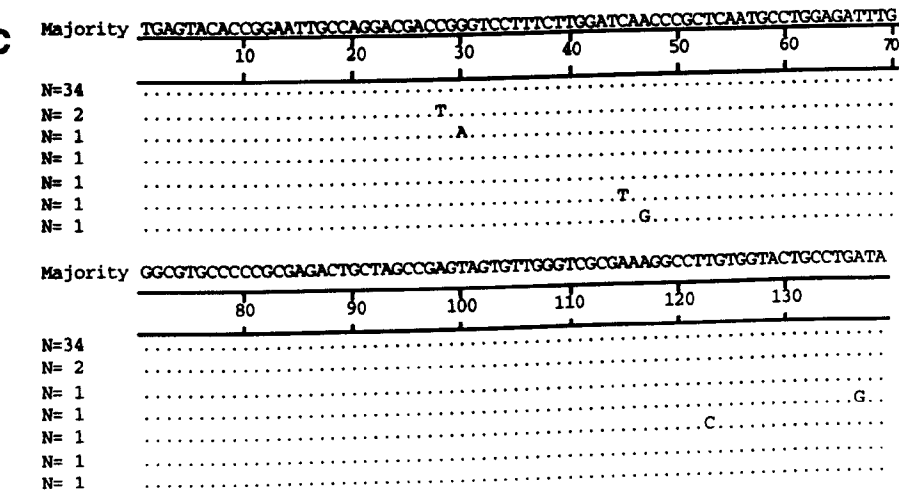
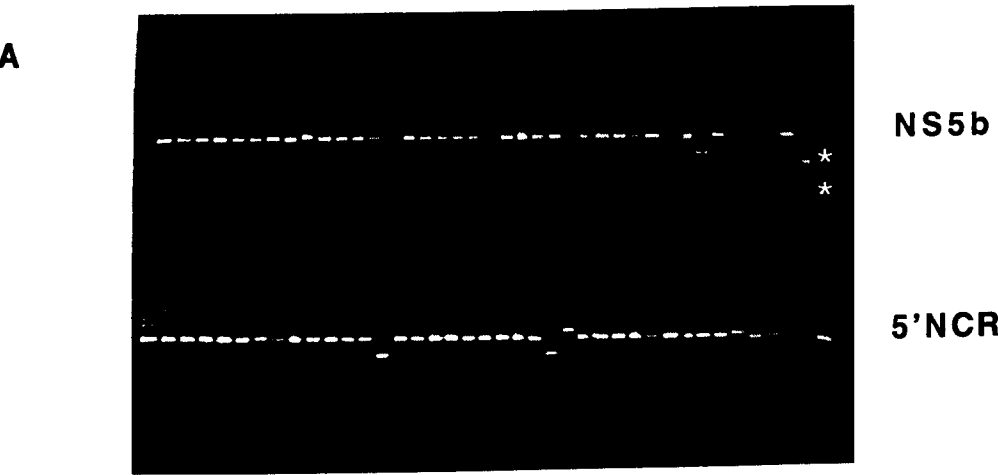
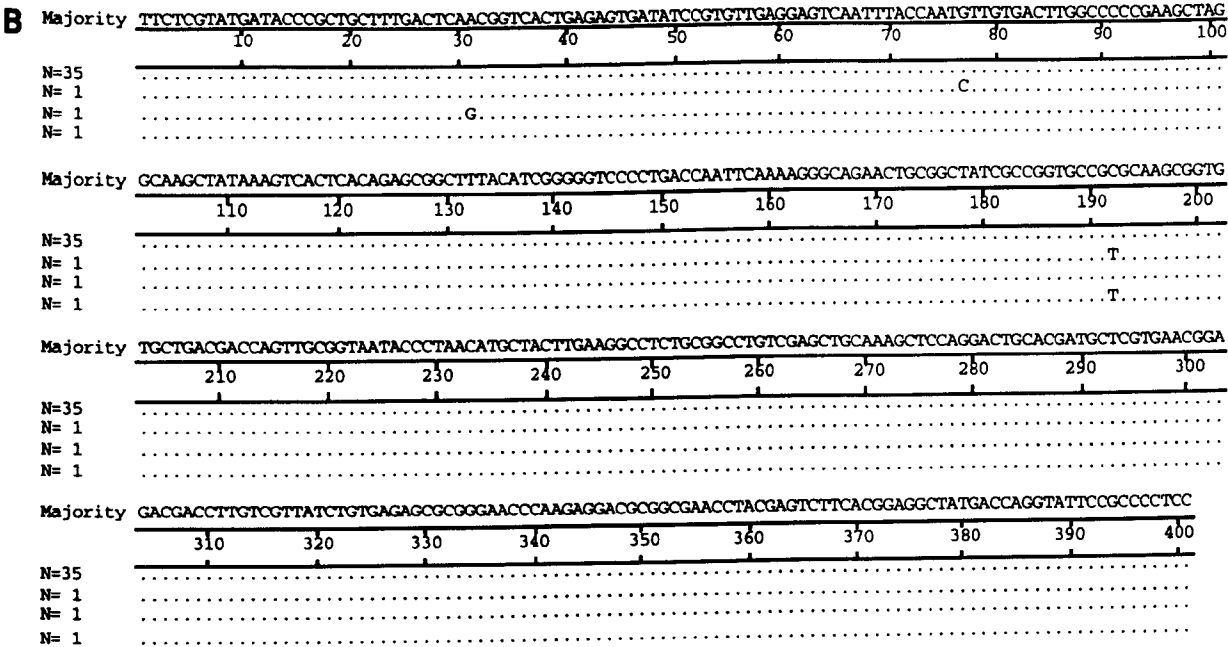


Figure 4.23: (A) DGGE analysis of clonotypes infecting patient IDU323. The gel shows the quasispecies distribution in the NS5b gene (upper bands) and the 5'NCR (lower bands). Clonotypes for which a nucleotide sequence was not obtained are indicated by *. (B) Nucleotide sequences of NS5b clonotypes infecting patient IDU323. (C) Nucleotide sequences of 5'NCR clonotypes infecting patient IDU323.

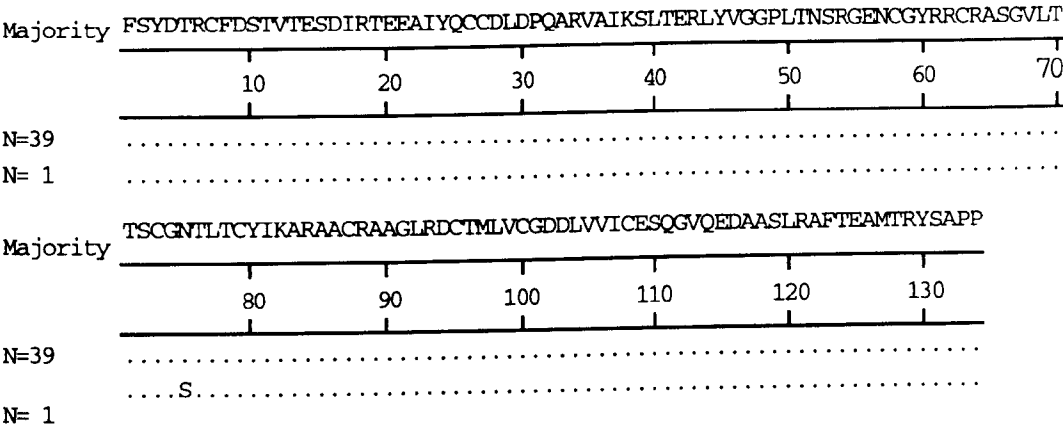


Figure 4.24: Amino acid sequences of the NS5b clonotypes infecting patient H069. Sequence identity is indicated by ‘.’.

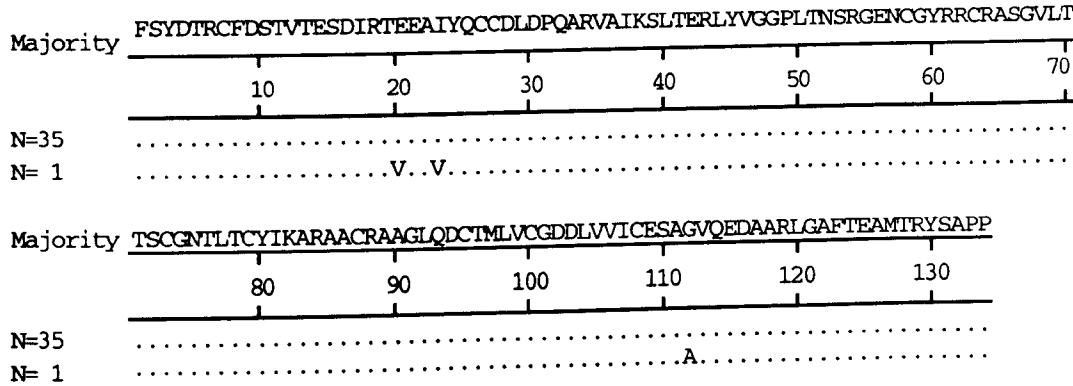


Figure 4.25: Amino acid sequences of the NS5b clonotypes infecting patient H071. Sequence identity is indicated by ‘.’.

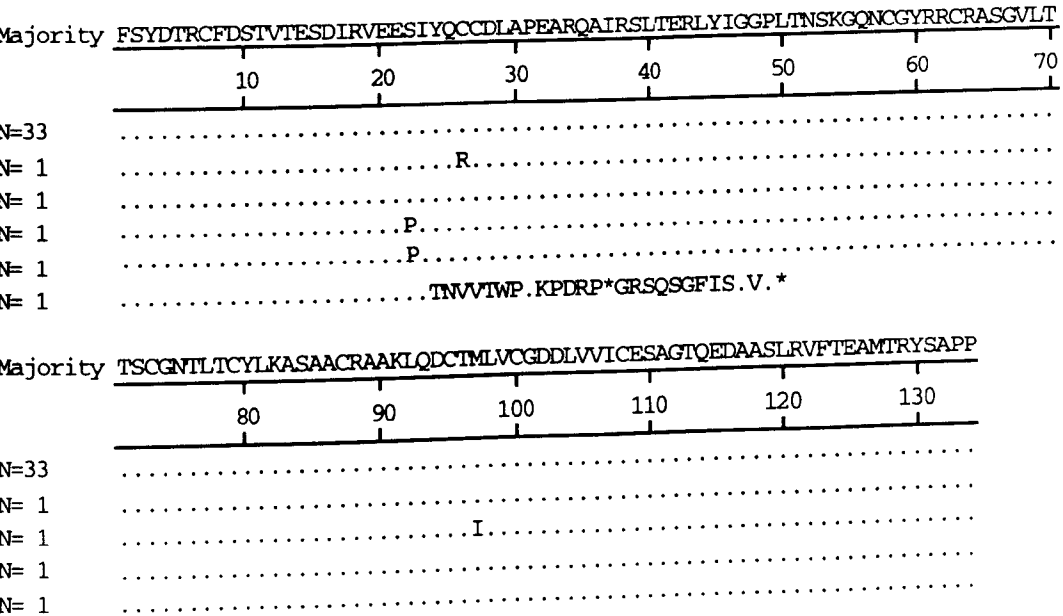


Figure 4.26: Amino acid sequences of the NS5b clonotypes infecting patient H075. Sequence identity is indicated by ‘.’. Stop codons are represented as *.

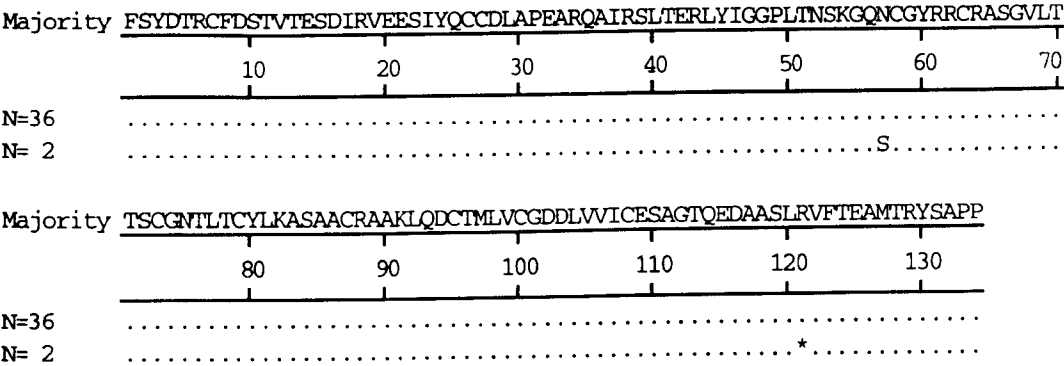


Figure 4.27: Amino acid sequences of the NS5b clonotypes infecting patient H858. Sequence identity is indicated by ‘.’. Stop codons are represented as ‘*’.

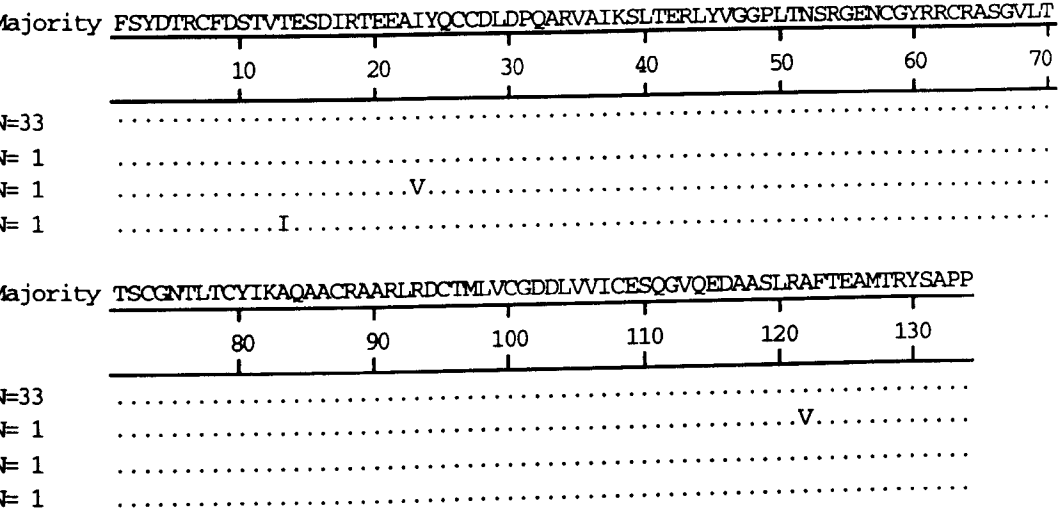


Figure 4.28: Amino acid sequences of the NS5b clonotypes infecting patient H865. Sequence identity is indicated by ‘.’.

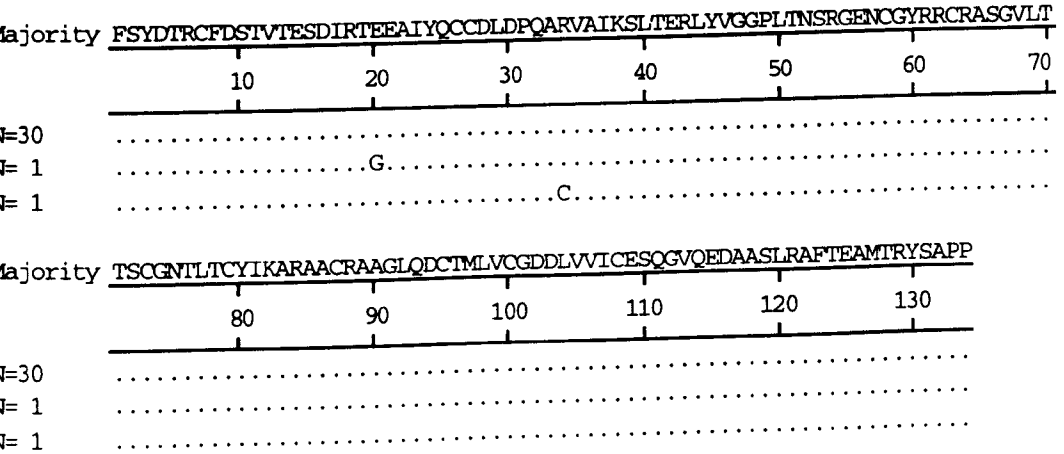


Figure 4.29: Amino acid sequences of the NS5b clonotypes infecting patient BD244. Sequence identity is indicated by ‘.’.

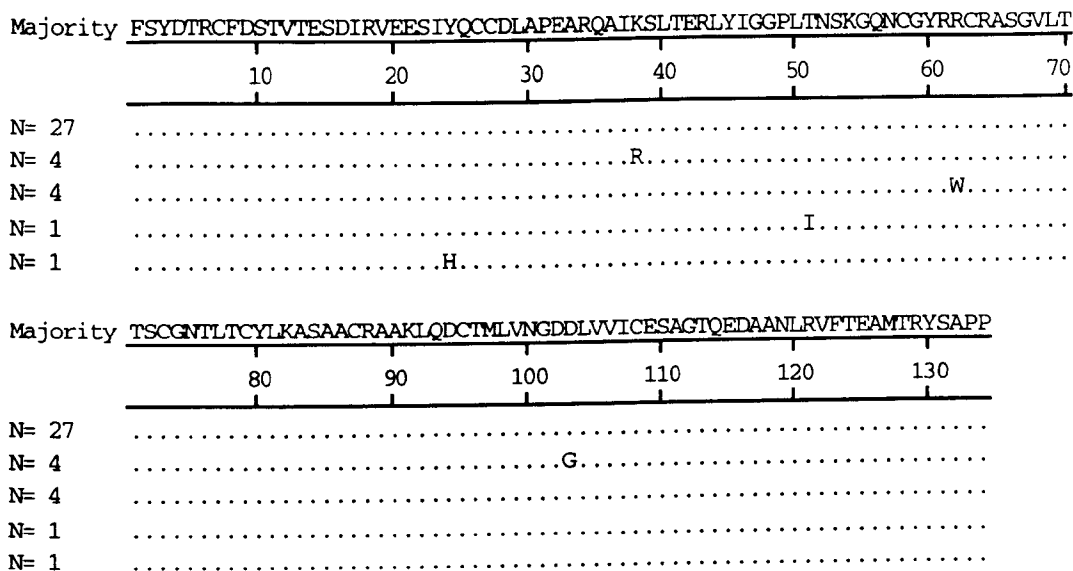


Figure 4.30: Amino acid sequences of the NS5b clonotypes infecting patient BD259. Sequence identity is indicated by ‘.’.

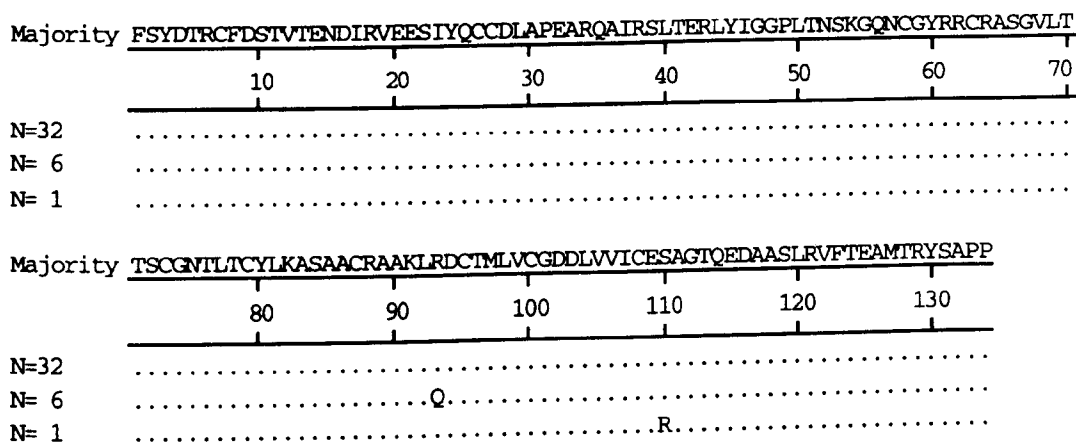


Figure 4.31: Amino acid sequences of the NS5b clonotypes infecting patient BD268. Sequence identity is indicated by ‘.’.

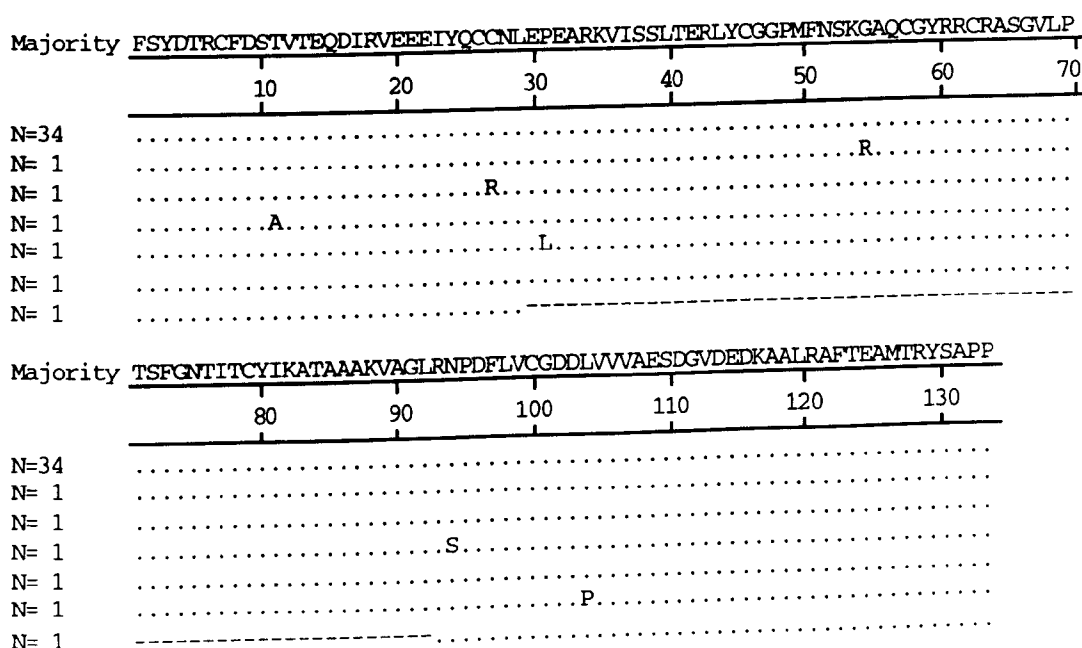


Figure 4.32: Amino acid sequences of the NS5b clonotypes infecting patient BD424. Sequence identity is indicated by ‘.’ and deletions are indicated by ‘-’.

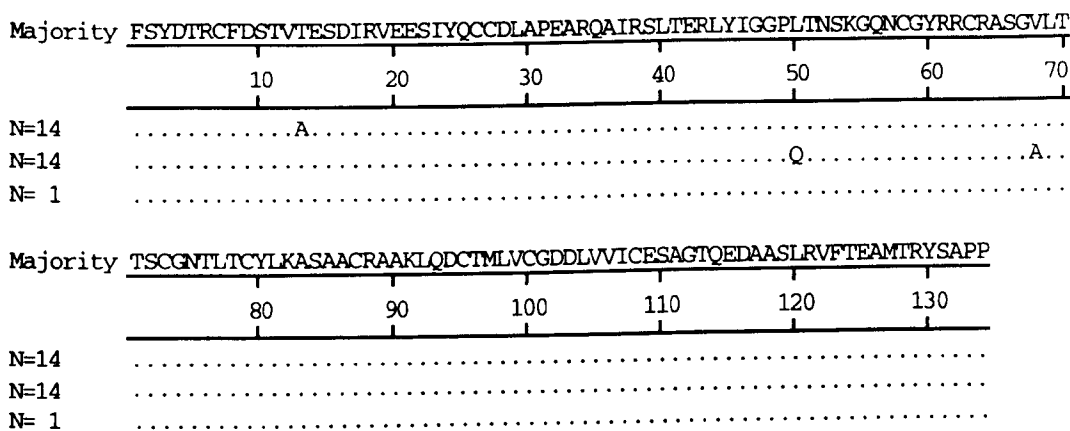


Figure 4.33: Amino acid sequences of the NS5b clonotypes infecting patient BD426. Sequence identity is indicated by '.'.

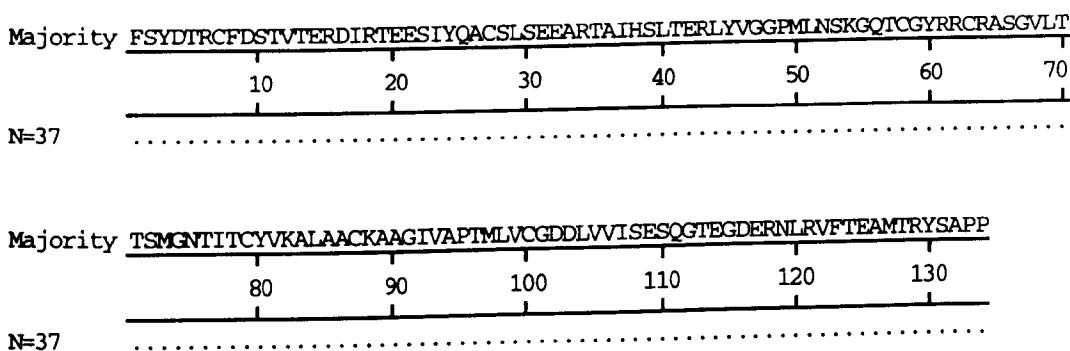


Figure 4.34: Amino acid sequences of the NS5b clonotypes infecting patient IDU189. Sequence identity is indicated by '.'.

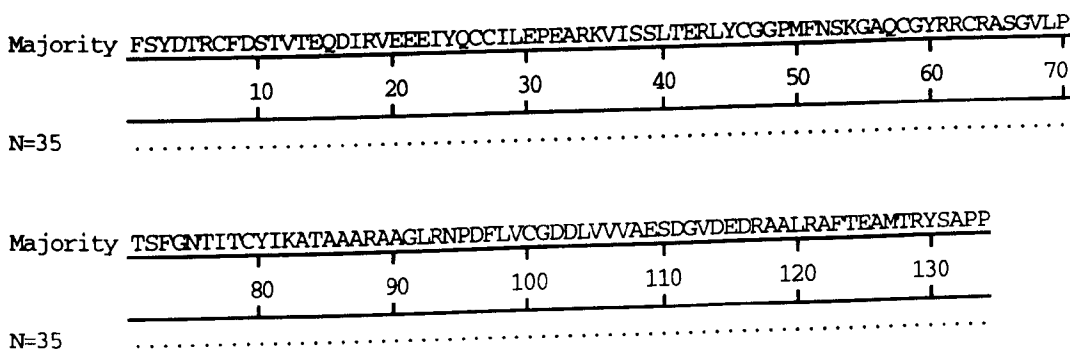


Figure 4.35: Amino acid sequences of the NS5b clonotypes infecting patient IDU191. Sequence identity is indicated by '.'.

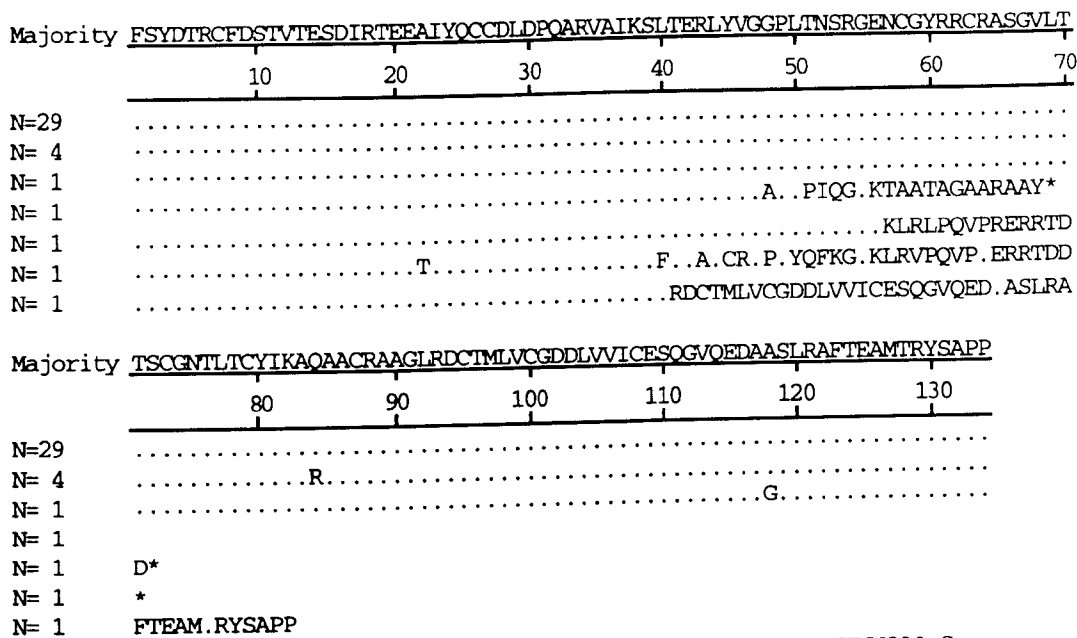


Figure 4.36: Amino acid sequences of the NS5b clonotypes infecting patient IDU230. Sequence identity is indicated by '.'. Stop codons are represented as *.

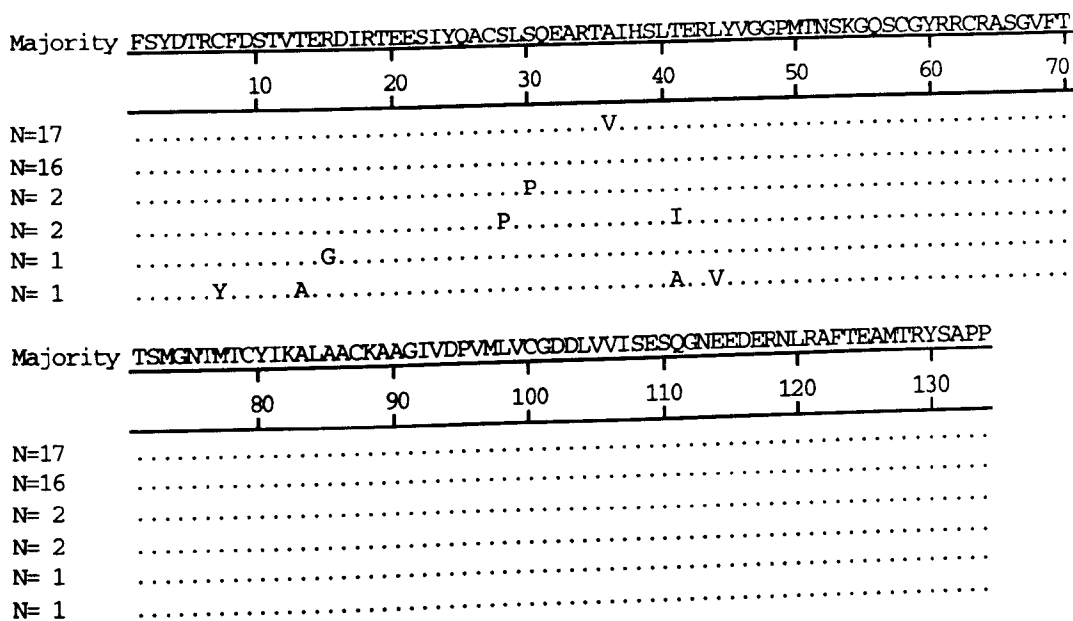


Figure 4.37: Amino acid sequences of the NS5b clonotypes infecting patient IDU240. Sequence identity is indicated by '.'.

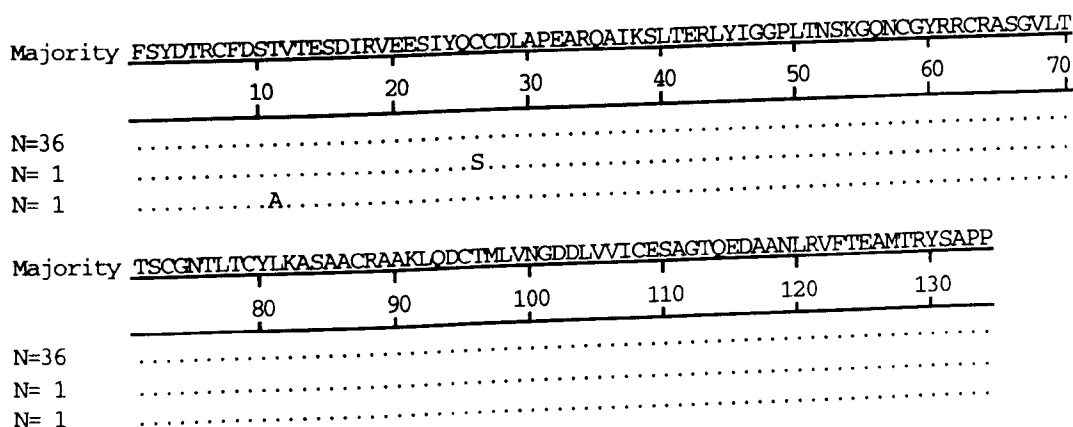


Figure 4.38: Amino acid sequences of the NS5b clonotypes infecting patient IDU323. Sequence identity is indicated by '.'.

4.3.2 Summary of diversity and sequence data

Tables 4.1 – 4.3 summarise the data obtained from DGGE analysis and nucleotide sequencing of HCV quasispecies infecting the 15 individuals in the study. Tables 4.4 – 4.6 show the average values of these characteristics for the 15 individuals, as a whole and within the three patient groups, for 5'NCR and NS5b nucleotide sequences and NS5b amino acid sequences.

Normalised Shannon entropy (S_n) (Pawlotsky et al. 1998b) is a measure of quasispecies complexity, incorporating the number of clonotypes, the frequency of each clonotype and the total number of colony PCR products analysed. S_n values can range from zero (when only one clonotype is present) to one (where each clonotype only occurs once), therefore higher values of S_n indicate higher quasispecies complexity. S_n values were calculated, as described in Section 2.6.15, for the 5'NCR nucleotide sequences, NS5b nucleotide sequences and NS5b amino acid sequences (Tables 4.1-4.3). S_n values ranged from 0.035 to 0.311 for the 5'NCR sequences; 0 to 0.502 for the NS5b sequences and from 0 to 0.334 for the NS5b amino acid sequences. The values are summarised in Tables 4.4 – 4.6. The hypothesis that the NS5b displayed greater sequence diversity than the 5'NCR was tested by comparing S_n values using the statistical methods described in Section 2.6.16. There was no evidence to support this hypothesis based on the difference between the two mean S_n values ($p > 0.05$).

The ranges and median values of S_n within each of the three patient groups are displayed in Tables 4.4 – 4.6. There was a significant difference ($p = 0.0358$) in S_n values for 5'NCR quasispecies between the three patient groups. Thus, haemophilia patients had significantly

lower diversity in this region than the blood donors and the IDUs. There was no significant difference in S_n values between any of the patient groups, for either NS5b quasispecies or NS5b amino acid sequences ($P > 0.05$).

The mean genetic distance (GD), the mean percentage nucleotide difference between clonotypes, is a measure of the quasispecies diversity of clonotypes. It can range from 0 to more than 100% (as the model of substitution used by CLUSTAL V assumes that 'multiple hits' can occur), the higher values representing higher diversity. Mean GDs were calculated, as described in Section 2.6.15, for 5'NCR nucleotide sequences, NS5b nucleotide sequences and NS5b amino acid sequences (Tables 4.1-4.3). Mean GD values (as percentage) ranged from 0.70 to 1.87 for 5'NCR nucleotide sequences, 0 to 1.93 for NS5b nucleotide sequences and 0 to 2.27 for NS5b amino acid sequences. These ranges and median values are displayed in Tables 4.4 – 4.6. There was a significant difference ($p = 0.029$) in mean GD values for 5'NCR quasispecies between the three patient groups. Thus, haemophilia patients had significantly lower diversity in this region than the blood donors and the IDUs. There was no significant difference in genetic difference values between any of the patient groups, for either NS5b quasispecies or NS5b amino acid sequences ($P > 0.05$).

The transition/transversion ratio (Ts/Tv) is the ratio of transitions to transversions. (A transition is the substitution of a purine for another purine ($G \leftrightarrow A$) or a pyrimidine for another pyrimidine ($C \leftrightarrow T$). A transversion is the substitution of a purine for a pyrimidine or vice versa.) Transitions tend to occur more readily than transversions and Ts/Tv ratios vary greatly from gene to gene. Ts/Tv ratios are useful for increasing the accuracy of many phylogenetic analysis methods, as it takes into account sequences that are heavily biased

towards particular types of changes (Clewley, 1998a). Ts/Tv ratios were calculated, as described in Section 2.6.14, for 5'NCR nucleotide sequences and NS5b nucleotide sequences. (Tables 4.1-4.2).

The ratio of synonymous/nonsynonymous (ds/dn) is a measure of how quickly a gene is changing and the type of selective pressure it may be under. Synonymous or "silent" mutations do not alter the amino acid sequences, but nonsynonymous mutations do, indicating the influence of selective pressure. Genetic drift, or neutral evolution, is signified by ds/dn ratios close to one, whereas negative selection against deleterious mutations gives rise to higher ds/dn ratios. A ds/dn ratio of less than one implies positive selection or that sequences are polymorphic (Clewley, 1998a). The ds/dn ratio was calculated, as described in Section 2.6.14, for NS5b nucleotide sequences only. Sequences with frame-shift mutations were excluded from this analysis. Ranges and median values of ds/dn within patient groups, and overall, are displayed in Table 4.5.

Patient	i	Sn	MeanGD(%)	Ts	Tv	Ts/Tv
H069	2	0.035	0.70	0	1	-
H071	2	0.047	0.70	1	0	-
H075	4	0.095	1.10	3	0	-
H858	3	0.083	0.70	1	1	1.0
H865	5	0.179	1.18	4	0	-
BD244	8	0.235	1.27	5	2	2.5
BD259	3	0.188	1.47	2	1	2.0
BD268	7	0.311	1.54	5	1	5.0
BD424	5	0.217	1.32	3	1	3.0
BD426	3	0.079	0.97	1	1	1.0
IDU189	9	0.264	1.78	9	2	4.5
IDU191	6	0.283	1.87	7	0	-
IDU230	7	0.209	1.40	4	2	2.0
IDU240	5	0.152	1.18	3	1	3.0
IDU323	7	0.202	1.27	6	0	-

Table 4.1: Number of clonotypes (i), normalised Shannon entropy (Sn), mean genetic distance (GD), Number of transistions (Ts), number of transversions (Tv) and transition/transversion ratio (Ts/Tv) for 5'NCR quasispecies of HCV infecting 15 individuals.

Patient	i	Sn	Mean GD (%)	Ts	Tv	Ts/Tv	ds	dn	ds/dn
H069	4	0.120	1.45	10	0	-	9	1	9.0
H071	2	0.036	0.80	1	2	0.5	0	3	-
H075	10	0.288	0.72	12	1	12.0	10	3	3.3
H858	9	0.466	1.42	17	0	-	15	2	7.5
H865	6	0.176	0.89	8	1	8.0	6	3	2.0
BD244	6	0.205	0.83	8	0	-	6	2	3.0
BD259	9	0.502	1.93	23	3	7.7	21	5	4.2
BD268	8	0.300	1.58	19	1	19.0	18	2	9.0
BD424	8	0.256	1.72	17	1	17.0	12	6	2.0
BD426	3	0.243	1.53	7	2	3.5	6	3	2.0
IDU189	4	0.103	0.53	4	0	-	4	0	-
IDU191	1	0.000	0.00	0	0	-	0	0	-
IDU230	9	0.282	1.06	14	2	7.0	14	3	4.7
IDU240	8	0.387	1.24	16	1	16.0	8	9	0.5
IDU323	4	0.099	0.43	2	1	2.0	1	2	0.9

Table 4.2: Number of clonotypes (i), normalised Shannon entropy (Sn), mean genetic distance (GD) (%), number of transistions (Ts), number of transversions (Tv), transition/transversion ratio (Ts/Tv), number of synonymous nucleotide changes (ds), number of non-synonymous nucleotide changes (dn) and synonymous/non-synonymous mutation ratio (ds/dn) for NS5b quasispecies of HCV infecting 15 individuals.

Patient	Sn	Mean GD (%)
H069	0.032	0.80
H071	0.036	2.30
H075	0.164	0.96
H858	0.057	0.80
H865	0.106	1.15
BD244	0.079	0.88
BD259	0.251	1.24
BD268	0.145	1.03
BD424	0.187	1.52
BD426	0.243	1.53
IDU189	0.000	0.00
IDU191	0.000	0.00
IDU230	0.252	1.03
IDU240	0.334	2.27
IDU323	0.067	1.03

Table 4.3: Normalised shannon entropy (Sn) and mean genetic distance (GD)(%) for translated NS5b quasispecies of HCV infecting 15 individuals.

Group	Range of Sn values (median)	Range of mean GD values (median)
Overall	0.035 – 0.311 (0.188)	0.70 – 1.87 (1.27)
Haemophilia patients	0.035 – 0.179 (0.083)	0.70 – 1.18 (0.7)
Blood donors	0.079 – 0.311 (0.217)	0.97 – 1.54 (1.32)
Intravenous drug users	0.152 – 0.283 (0.209)	1.18 – 1.87 (1.40)

Table 4.4: Ranges and median values for normalised Shannon entropy (Sn) and genetic distance (GD) (%) for 5'NCR quasispecies of HCV infecting the 15 individuals as a whole and the three patient groups.

Group	Range of Sn values (median)	Range of mean GD values (median)	Range of ds/dn Values (median)
Overall	0 – 0.502 (0.243)	0 – 1.93 (1.06)	0.5 – 9.0 (3.2)
Haemophilia patients	0.036 – 0.466 (0.176)	0.72 – 1.45 (0.89)	2.0 – 9.0 (5.4)
Blood donors	0.205 – 0.502 (0.256)	0.83 – 1.93 (1.58)	2.0 – 9.0 (3.00)
Intravenous drug users	0 – 0.387 (0.103)	0 – 1.24 (0.53)	0.5 – 4.7 (0.9)

Table 4.5: Ranges and median values for normalised Shannon entropy (Sn), genetic distance (GD) (%) and synonymous/non-synonymous mutation ratio (ds/dn) for NS5b quasispecies of HCV infecting the 15 individuals as a whole and the three patient groups.

Group	Mean Sn	Mean GD (%)
Overall	0 – 0.334 (0.106)	0 – 2.30 (1.03)
Haemophilia patients	0.032 – 0.164 (0.057)	0.8 – 2.30 (0.96)
Blood donors	0.079 – 0.251 (0.187)	0.88 – 1.53 (1.24)
Intravenous drug users	0 – 0.334 (0.067)	0 – 2.27 (1.03)

Table 4.6: Ranges and median values for normalised Shannon entropy (Sn) and genetic distance (GD) (%) for translated NS5b quasispecies of HCV infecting the 15 individuals as a whole and the three patient groups.

4.3.3 Genotyping clonotypes by phylogenetic analysis

Genotypes had been assigned to the HCV infecting the 15 individuals in this study by direct RFLP analysis of PCR products amplified from the 5'NCR (see Chapter 3). Table 4.7 shows the genotype assigned to each of the 'consensus' HCV 5'NCR sequences by this method and the genotypes assigned using phylogenetic analysis of the 5'NCR and the NS5b genes.

The nucleotide sequences of the 5'NCR and NS5b clonotypes from each of the 15 patients were analysed using the phylogenetic methods described in Section 2.6.14. Sequences representing each of the genotypes and the main subtypes in both genomic regions, obtained from the Genbank database, were analysed alongside the clonotypes. Clonotypes were assigned genotypes according to their clustering pattern with Genbank sequences (Simmonds et al. 1993; Mellor et al. 1995).

Patient	RFLP analysis of the 5'NCR	Sequence analysis of the 5'NCR	Sequence analysis of the NS5b
H069	1a	1b	1a
H071	1a	1b	1a
H075	1b	1b	1b
H858	1b	1b	1b
H865	1a	1b	1a
BD244	1a	1a	1a
BD259	1b	1b	1b
BD268	1b	1b	1b
BD424	3a	3a	3a
BD426	1b	1b	1b
IDU189	2a	1a	2a
IDU191	3a	3a	3a
IDU230	1a	1b	1a
IDU240	2b	2b	2b
IDU323	1b	1b	1b

Table 4.7: The genotypes of HCV infecting the 15 individuals in this study determined by RFLP analysis of whole PCR product derived from the 5'NCR and sequence analysis of clones derived from the 5'NCR and the NS5b.

Figures 4.39 – 4.50 show the dendrograms produced using the CLUSTAL algorithm in the MEGALIGN program and DNADIST and FITCH (viewed in TREEVIEW) from the PHYLIP suite of programs. The dendrograms show similar topologies, allowing confidence in the use of clustering as the basis of genotype assignment. Sequences were analysed as three patient groups because the whole data set was too large to be analysed by the programs. A genotype was assigned to each set of clones from each individual, using the trees generated with sequences from both genomic regions, and are displayed in Table 4.7. None of the individuals were infected with more than one genotype or subtype. The genotypes assigned by RFLP analysis of the 5'NCR always agreed with those assigned by sequence analysis of the NS5b but did not always agree with the genotypes assigned by sequence analysis of the 5'NCR.

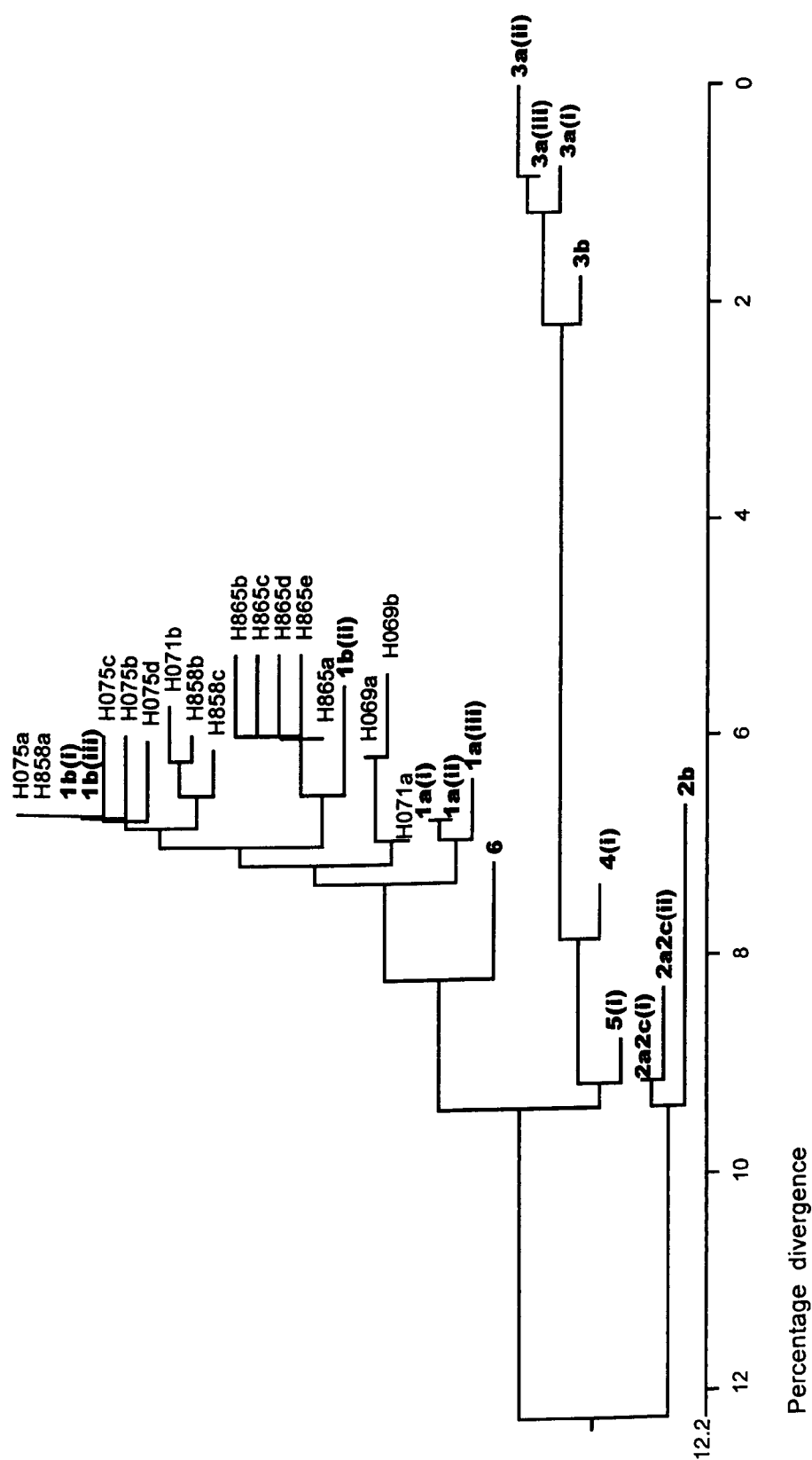


Figure 4.39: Unrooted tree showing how 5'NCR clonotype sequences infecting the haemophilia patients cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the CLUSTALV algorithm in the MEGALIGN program.



Figure 4.40: Unrooted tree showing how 5'NCR clonotype sequences infecting the haemophilia patients cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

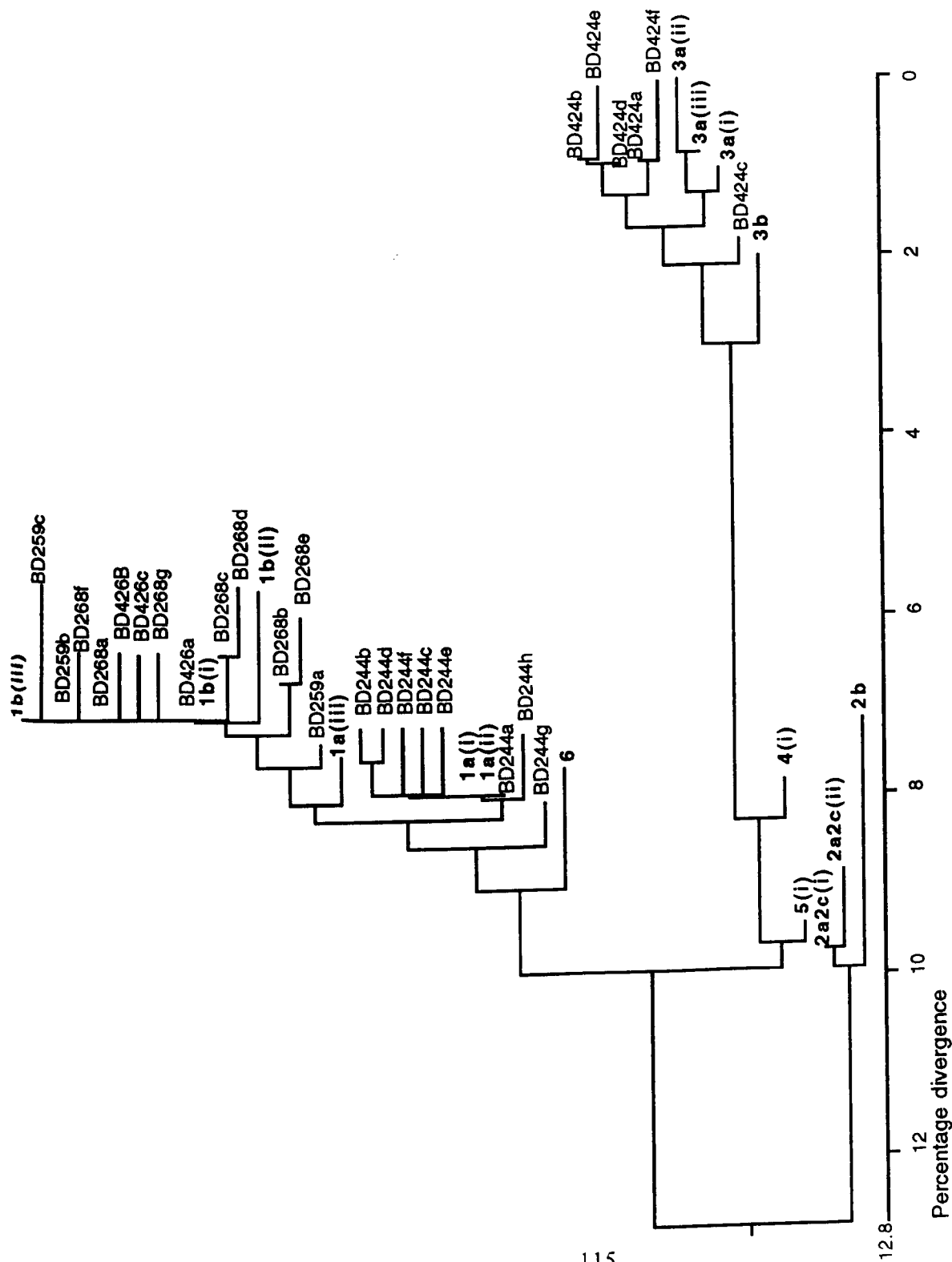


Figure 4.41: Unrooted tree showing how 5'NCR clonotype sequences infecting the blood donors cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the CLUSTALV algorithm in the MEGALIGN program.



Figure 4.42: Unrooted tree showing how 5'NCR clonotype sequences infecting the blood donors cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

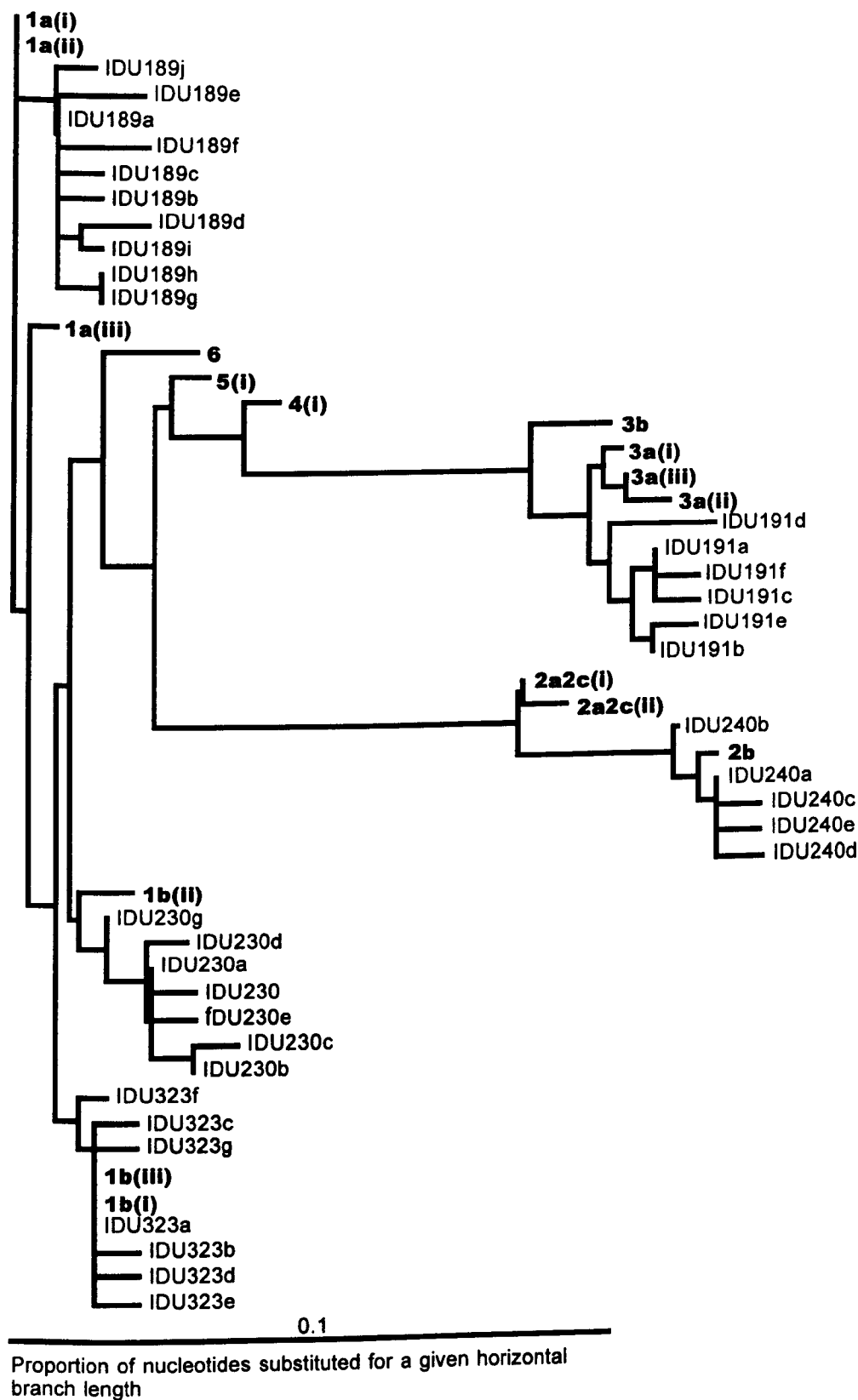


Figure 4.44: Unrooted tree showing how 5'NCR clonotype sequences infecting the IDUs cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

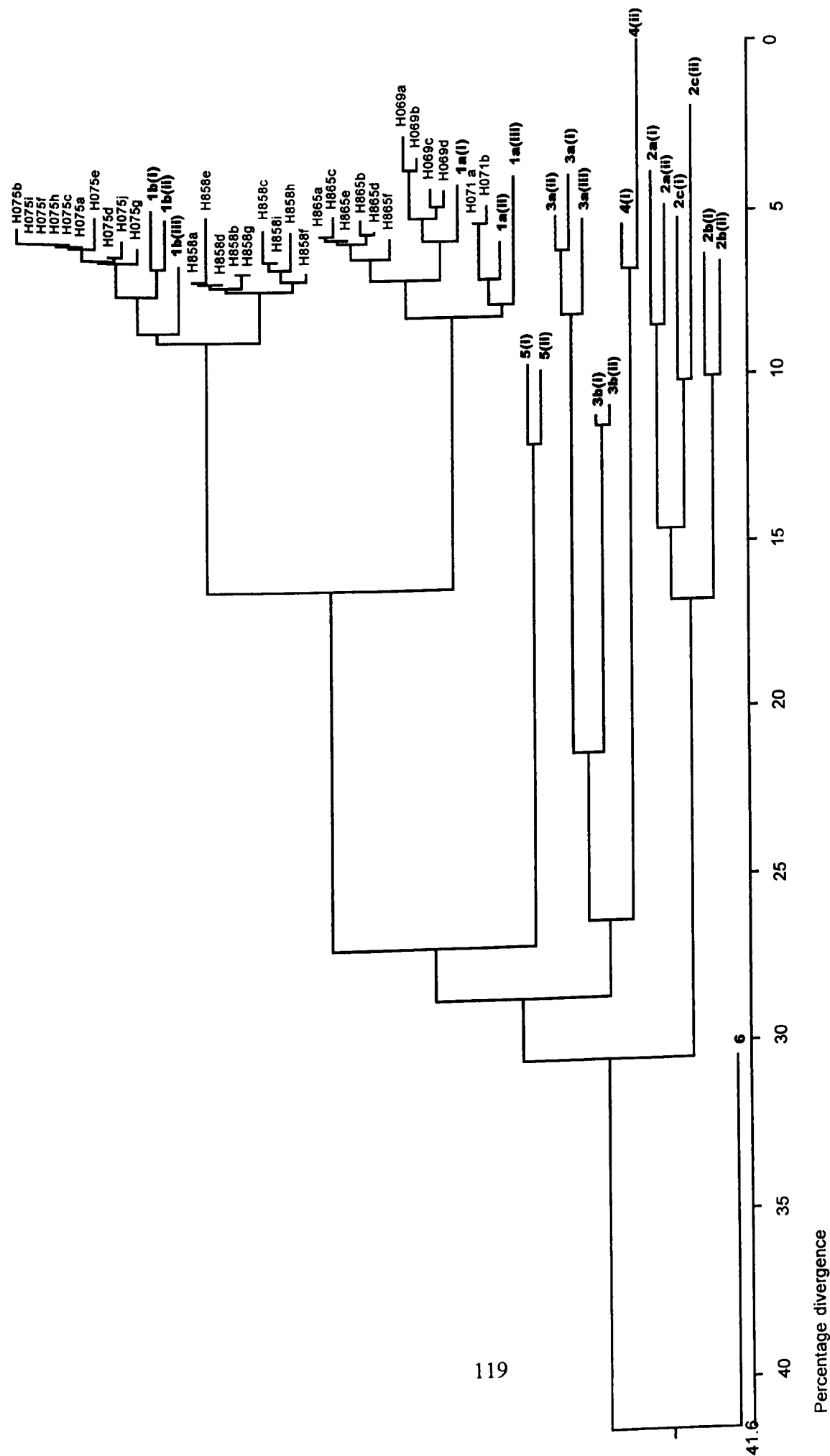


Figure 4.45: Unrooted tree showing how NS5b clonotype sequences infecting the haemophilia patients cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the CLUSTALV algorithm in the MEGALIGN program.

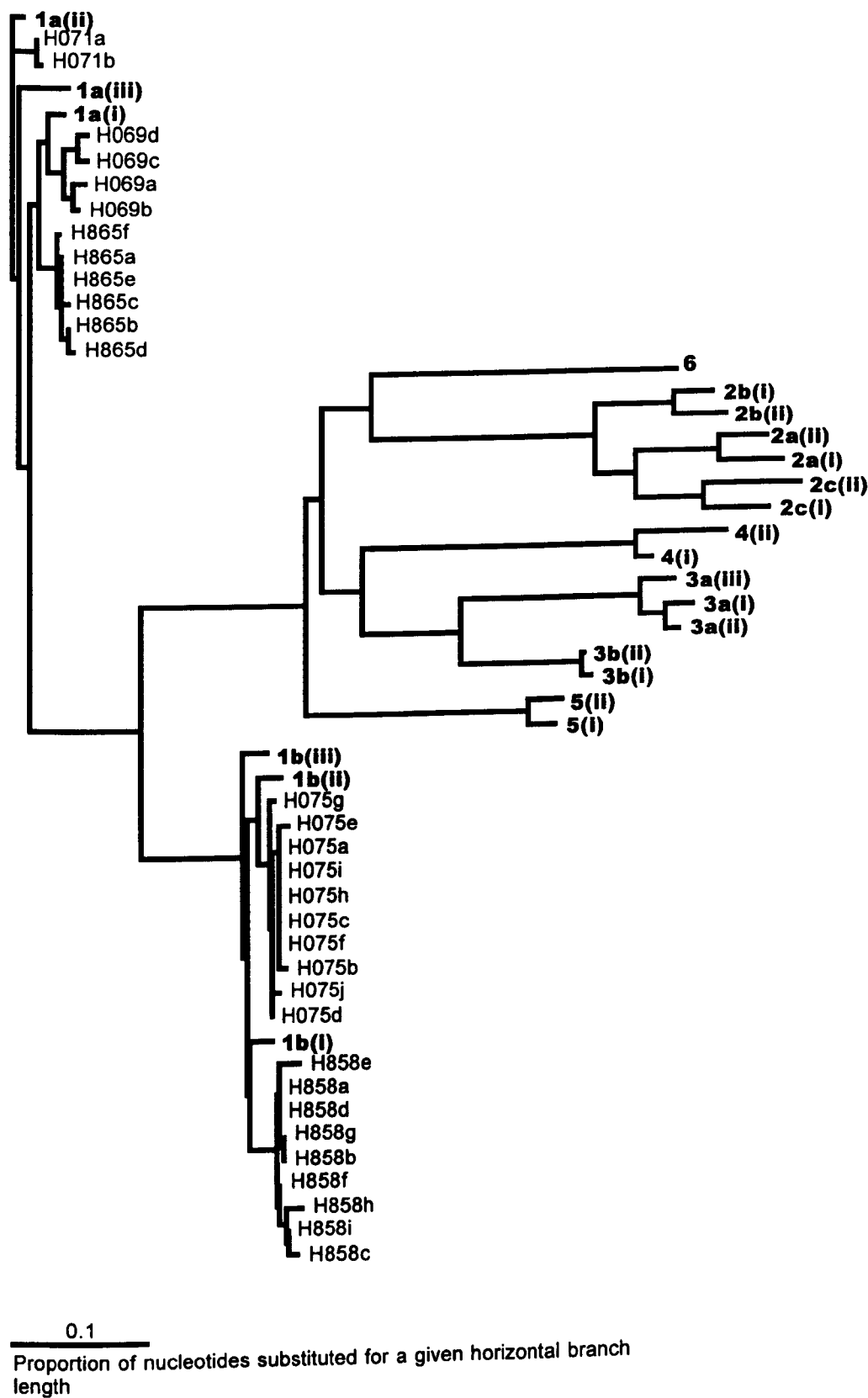


Figure 4.46: Unrooted tree showing how NS5b clonotype sequences infecting the haemophilia patients cluster with sequences of known genotype obtained from the Genbank database (in **bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

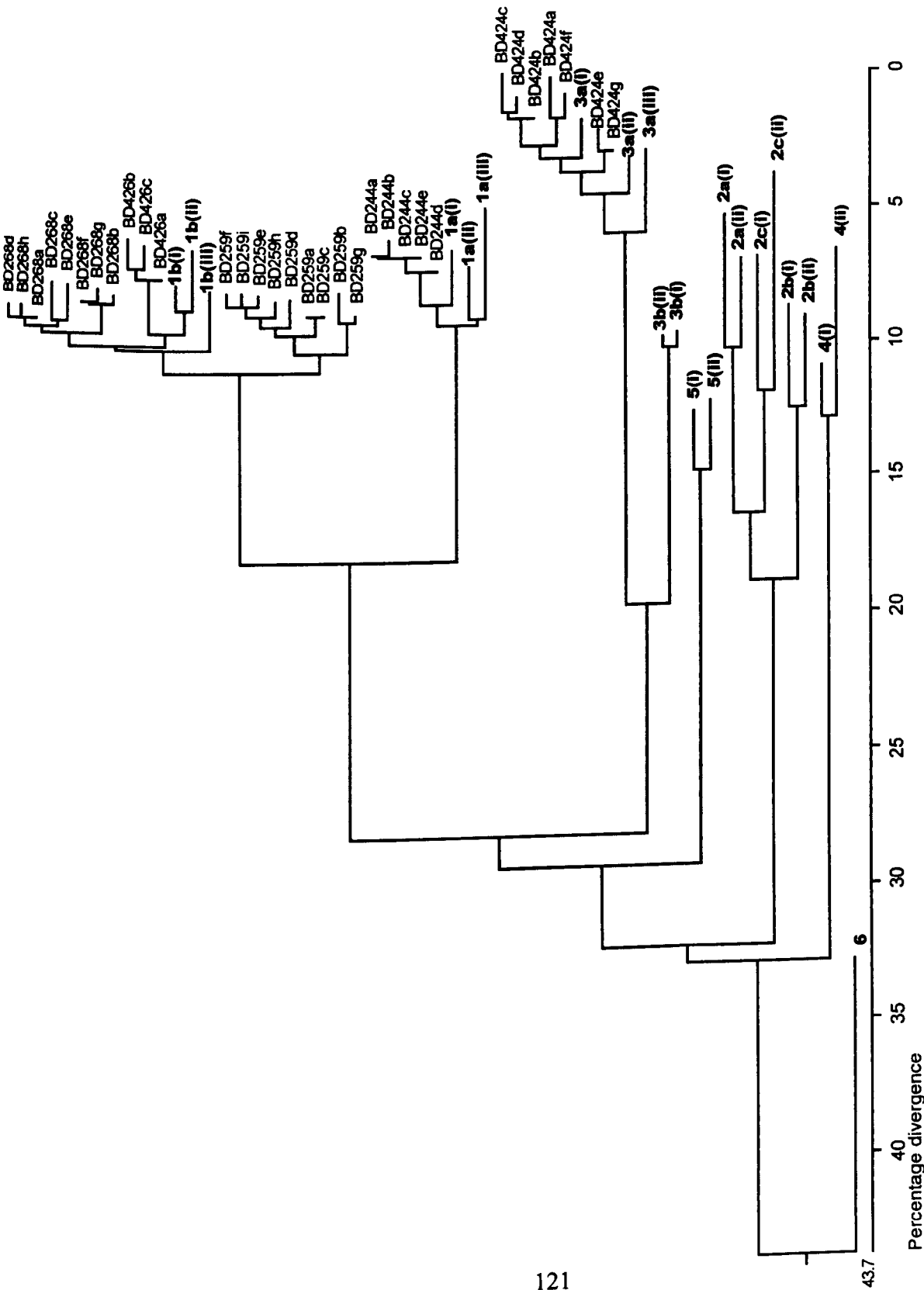


Figure 4.47: Unrooted tree showing how NS5b clonotype sequences infecting the blood donors cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the CLUSTAL V algorithm in the MEGALIGN program.

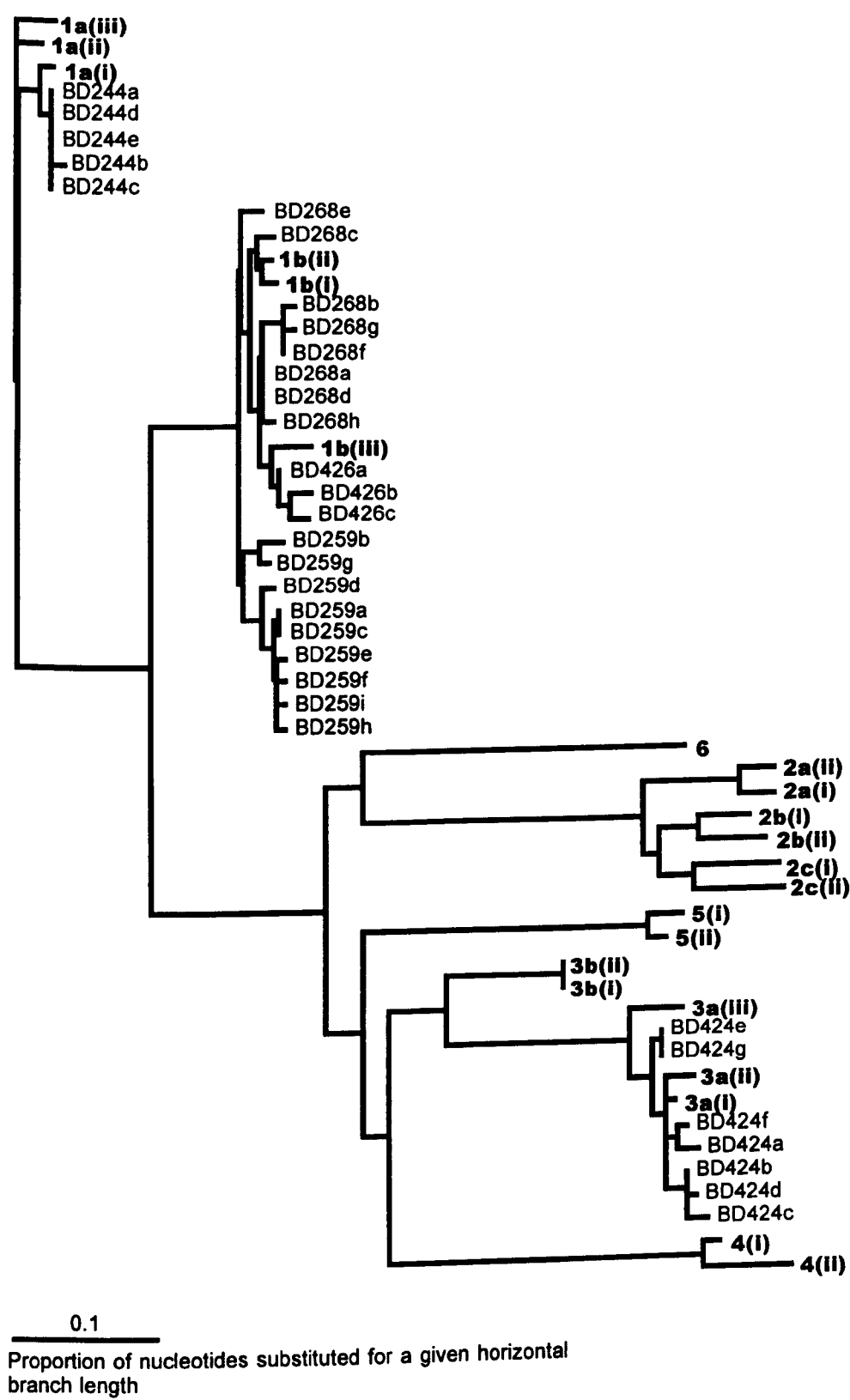


Figure 4.48: Unrooted tree showing how NS5b clonotype sequences infecting the blood donors cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

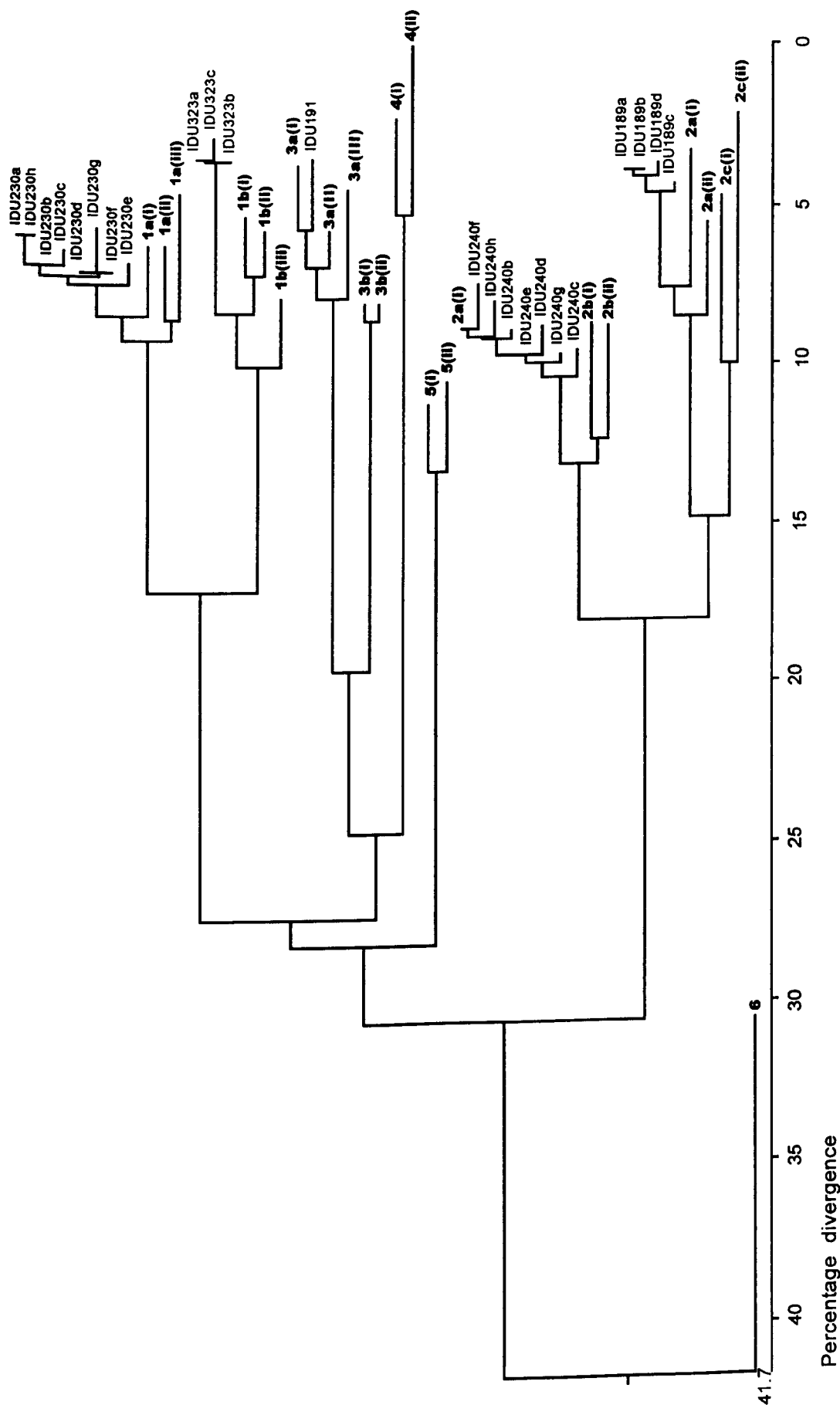


Figure 4.49: Unrooted tree showing how NS5b clonotype sequences infecting the IDUs cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the CLUSTALV algorithm in the MEGALIGN program.

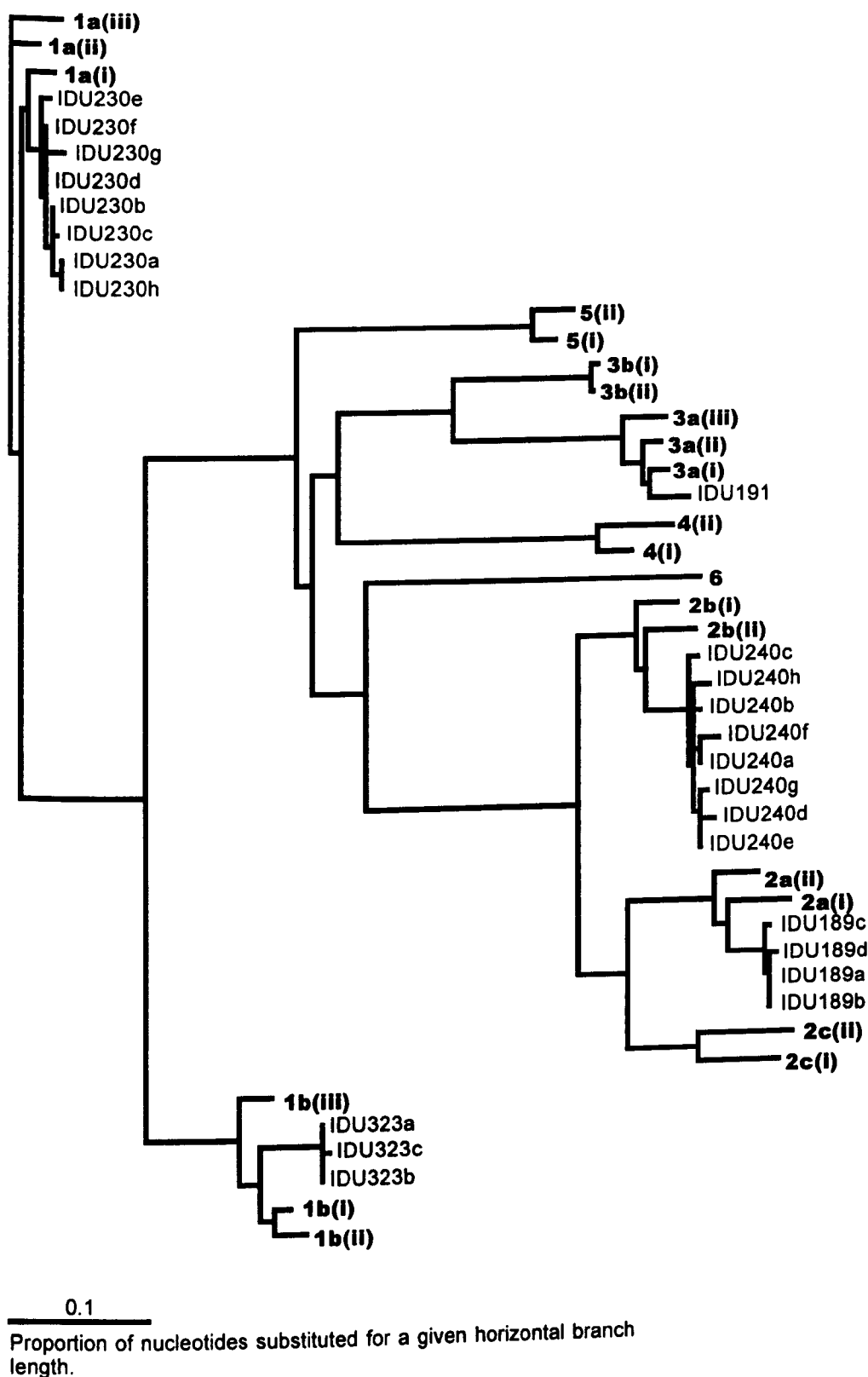


Figure 4.50: Unrooted tree showing how NS5b clonotype sequences infecting the IDUs cluster with sequences of known genotype obtained from the Genbank database (**in bold type**). Clones from individuals are shown in plain type. Dendrogram was produced using the FITCH program in the PHYLIP package.

4.4 Discussion

4.4.1 Quasispecies complexity and diversity

I have described the complexity and diversity of the HCV quasispecies infecting 15 individuals. By developing a DGGE based technique I was able to screen a large number (~80) of species from each individual. Evaluating quasispecies complexity by DGGE analysis was a very rapid procedure. However, to measure the diversity of the quasispecies population it was necessary to determine nucleotide sequences for each clonotype identified by DGGE. The sequencing done here, although restricted, was already both labour intensive and expensive. This highlighted how difficult it would have been to sequence all 80 clones obtained from each of the 15 patients without using a prior screening method such as DGGE.

Figures 4.9 – 4.23 illustrate the quasispecies complexity and diversity of HCV infecting 15 individuals. Data relating to sequence heterogeneity could be summarised by calculating Shannon Entropy values, a measure of quasispecies complexity, and mean genetic distances, a measure of quasispecies diversity (Tables 4.1 – 4.6). Genetic complexity was generally higher for the NS5b gene than for the 5'NCR. This is probably due to the constraints placed on the number of possible mutations in the 5'NCR in coding for the IRES structure (Smith et al. 1995). The NS5b gene, which codes for the RNA polymerase, appears to tolerate more mutations. However, some of these alter the reading frame of the gene resulting in amino acid sequences that probably do not represent a functional RNA polymerase protein (See Figures 4.9 - 4.38). Martell et al. (1992) described a high proportion of mutations that gave rise to defective HCV genomes and concluded that most

circulating viral particles carry defective genomes. These viral particles are termed 'defective interfering particles' and they require competent virus to replicate, and they replicate in preference to wild-type virus. Defective interfering particles interfere with competent virus production by competing for host cell components of replication, hence they can modulate levels of wild type virus and may protect tissue against acute cell death (Perrault, 1981; Barrett and Dimmock, 1986; Holland, 1991). The production of defective particles may provide an important mechanism, in addition to genetic diversity, by which HCV establishes chronic infection (Martell et al. 1992; Higashi et al. 1993). Higashi et al. (1993), in particular, observed a striking fluctuation in the proportion of defective HCV over time in chronically infected patients. In the study described in this chapter, one individual, IDU230, carried HCV genomes of which 11% contained apparently defective NS5b genes (See Figure 4.37). Such findings may reflect temporal fluctuations in the proportion of circulating defective viral particles.

Genetic complexity varied between individuals in both the 5'NCR and the NS5b. If the host immune response does influence the pace of viral evolution, this may reflect differences in the nature of the host immune response between individuals and over time (Farci et al. 1997; Ni et al. 1997). Treatment with interferon may also influence the quasispecies distribution (Nagasaka et al. 1996; Gonzalez-Peralta et al. 1997). However, none of the haemophilia patients were known to have received interferon treatment. HCV infections in the blood donors and IDUs were identified via anonymous screening and so it is unlikely that any of these patients were receiving any antiviral therapy.

4.4.2 Mode of acquisition and diversity

In addition to investigating the variation in quasispecies complexity and diversity between individuals, the design of this study allowed the comparison of these characteristics between individuals with different modes of acquisition of HCV. Inoculation with HCV in relatively large volumes of infectious material (e.g. blood products) would introduce a greater variety of quasispecies into the newly infected individual than a smaller inoculum (e.g. via contaminated needles) (Pawlotsky, 1998a). The complexity and diversity of quasispecies in chronic HCV infection may be influenced by the diversity of the infectious dose. If this were the case I would have expected to see a more complex and diverse quasispecies distribution among the haemophilia patients than among the other two patients groups. Individuals with a high risk of multiple exposures to HCV may also be expected to display high levels of quasispecies complexity and diversity (Toyoda et al. 1996). However, neither the haemophilia patient group nor the injecting drug user group displayed more complex or diverse quasispecies distributions.

4.4.3 Ratio of synonymous to non-synonymous changes

The ds/dn ratio was greater than one in most cases, (See Table 4.2), indicating that the NS5b gene is under negative selective pressure (ie deleterious mutations were being selected against). The exceptions were patients IDU240 and IDU323, for whom values were less than one suggesting positive selective pressure. The IDU group showed a lower median value for the ds/dn ratio, (Table 4.5), indicating that the NS5b gene of HCV infecting IDUs was under less pressure. Whether there is a biological basis for this observation is unknown. As the total number of nucleotides and mutations was relatively small, the number of synonymous and non-synonymous mutations were calculated by

simply counting nucleotide changes in the alignments. However, this method does not allow for the possibility of back-mutations or multiple mutations at a single position. When large data sets are studied and the sequence variation is greater, a model of substitution, such as Jukes-Cantor, can correct for these factors (Kumar et al. 1993).

4.4.4 Genotype and subtypes

The HCV infecting the 15 patients in this study had been assigned a genotype using direct RFLP analysis of an amplified fragment derived from the 5'NCR (See Sections 2.1 and 2.2) as part of the large genotyping study described in Chapter 3. The current Chapter shows nucleotide sequence data and derived phylogenetic analysis of all of the clonotypes infecting the 15 individuals, this being done to confirm the HCV genotypes assigned. It was not possible to analyse all of the clonotypes from the 15 individuals simultaneously, since there were too much data for the phylogenetic programs to process. I chose to analyse clonotypes from each of the three patient groups separately, along with sequences obtained from the Genbank database that were representative of all six genotypes and the main subtypes. By constructing the dendrograms displayed in Figures 4.39 - 4.50, using two phylogenetic methods, CLUSTAL V and DNADIST with FITCH (using TREEVIEW to display the output), I was able to assign genotypes and subtypes to all of the clonotypes infecting the 15 individuals. Each of the dendrograms produced with DNADIST and FITCH was the consensus of ten different trees arising from different input orders of sequences into the program. Thus, it is more rigorous than CLUSTAL V. Although both methods used to construct the dendrograms were similar, the fact that the dendrograms show similar topologies allows one to have greater confidence that the way in which sequences cluster is an accurate description of their relatedness.

The dendrograms (Figures 4.39 - 4.50) revealed several features. First, while genotypes assigned by sequence analysis of NS5b clonotypes were in complete agreement with those previously assigned by PCR-RFLP, phylogenetic analysis of 5'NCR sequences produced discrepant results for five of the individuals: H069, H071, H865, IDU189 and IDU 230 (Table 4.7). The majority of discrepancies were due to mis-typing of 1a as 1b, or vice versa. It has been reported that around 15% of type 1a may be mistyped as 1b when the genotyping method is based on the 5'NCR (Davidson et al. 1995). It is clear from the 5'NCR dendrograms (Figures 4.39 - 4.50) that there was simply not enough variation in the 5'NCR to allow discrimination between subtypes. The sequence polymorphisms that allow genotyping by RFLP analysis of the 5'NCR were lost in the 'background noise' of other single base mutations that occurred in this region. Second, phylogenetic analysis of NS5b clonotypes produced dendrograms that clearly distinguished between genotypes and subtypes; genetic distances between the various types were much greater in this region than in the 5'NCR. The agreement between genotype assignment with NS5b sequence analysis and the RFLP analysis of the 5'NCR therefore shows the PCR-RFLP assay to be a relatively robust method for genotyping.

4.4.5 Infection with multiple genotypes or subtypes

None of the 15 individuals were infected with more than one genotype or subtype. I have already observed from the findings in Chapter 3 that the prevalence of mixed genotype/subtype infections detected by PCR-RFLP was low. One possibility entertained to explain this low prevalence is that other genotypes or subtypes may have been present at very low concentrations, undetectable by direct PCR analysis (see Section 3.4.2). However, the study in the present chapter again revealed that mixed genotypes or subtypes were absent. I re-investigated the sensitivity of the PCR-RFLP based genotyping system

employed in Chapter 3 to verify if genotypes or subtypes that could be circulating as minor variants can indeed be detected by the PCR-RFLP method. Figure 4.51 shows that it detects a genotype or subtype co-existing with another type at a ratio of about 1 in 50. Since the cloning approach described in this chapter evaluated about 40 clones per specimen, the sensitivity limits of PCR-RFLP and PCR-cloning are equivalent. Thus, using procedures that are able to detect minor variants comprising as low as 5% in a mixture, shows mixed infection to be rare.

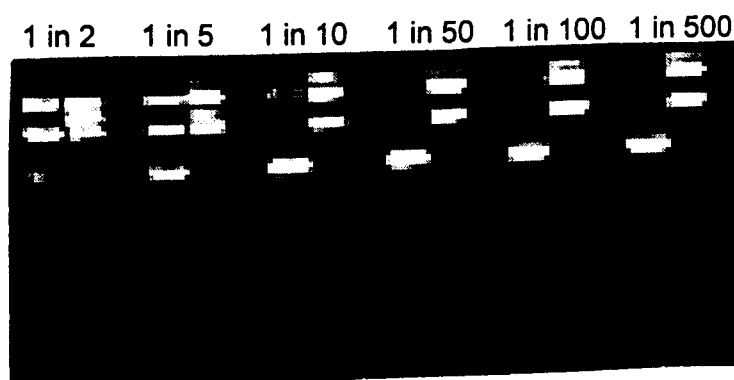


Figure 4.51: Genotyping PCR products, using the PCR-RFLP assay described in Chapter 3, amplified from sera containing HCV types 1a and 3a. The digests were loaded as described in Figure 3.3. The viral load for each serum sample was determined using the Roche Monitor 2.0 kit (Roche Diagnostic systems, Branchburg, USA). The proportion of genotype 3a serum diluted in genotype 1a serum is shown above the RFLP pattern for each sample.

Why is infection with multiple subtypes and genotypes rare? The quasispecies model provides one explanation. The competitive exclusion principle of population biology states that when two variants exist in the same environment, even when they have equal relative fitness, one variant will eventually prevail. A highly advantageous mutation in one or a deleterious mutation in the other will lead to the sudden exclusion of the least fit variant (Domingo et al. 1996). Relating this to superinfection with HCV, the fittest component of a

newly acquired quasispecies will compete with the existing dominant strain until one or the other prevails. If these competing strains are of different genotypes or subtypes then either the existing genotype will continue to prevail or the new genotype will overtake it, resulting in a change in the dominant circulating genotype. Several studies of superinfection, in chimpanzees and humans, support this model. Farci et al. (1992) demonstrated that while chronically infected chimpanzees displayed bouts of acute hepatitis when challenged with heterologous strains of HCV, no change in the prevailing genotype was observed. Similar findings were made by Okamoto et al. (1994) who observed no takeover in one infected chimpanzee inoculated with HCV of a different genotype. However, in two other animals, superinfection resulted in a change in the circulating genotype. Studies in humans also show that superinfection can occur and that the subsequent genotype detected will either remain the same or be different (Kao et al. 1993; Jarvis et al. 1994). Multiple genotypes or subtypes can co-circulate transiently, explaining the low incidence of detection of multiple genotypes and subtypes.

The superinfecting virus faces a disadvantage, as existing quasispecies populations tend to suppress minor variants, even if they are of superior fitness (Domingo et al. 1995). Deliberate seeding of variants into viral populations of lower fitness has shown that such variants do not always rise to dominance (de la Torre and Holland, 1990). The implication of this is that the route to HCV vaccination lies in the use of live, attenuated HCV rather than inactivated or subunit vaccines. However, the potential risk of reversion to virulence may be too high for this to be a realistic strategy.

It is possible that different types could persist at different sites within an infected individual, such as liver cells or PBMCs, as these represent environmental niches. Changes

in environmental conditions during chronic infection, such as antiviral treatment or co-infection with HIV, may favour the selection of a new dominant strain and a turnover in the prevailing genotype (Farci et al. 1992; Jarvis et al. 1994; Pujol et al. 1998). However, the distinction between reactivation, if such a process actually exists for HCV, and superinfection is difficult to establish.

4.4.6 Sources of artefactual error

Artefactual sources of variation can influence studies of viral diversity. One arises from misincorporation errors occurring during PCR. The enzyme most commonly used in the PCR amplification of DNA is *Taq* DNA polymerase. Most studies estimate that the misincorporation rate of *Taq* is $0.2\text{--}0.3 \times 10^{-4}$ errors per base pair per cycle of PCR (Eckert and Kunkel, 1991; Smith et al. 1997; Bracho et al. 1998). If this figure is applied to a typical nested PCR amplification of a 200 base pair fragment, after 30 “effective” cycles, 7% of the amplicons would be expected to contain a misincorporation error. The number of “effective” cycles is approximately 60% of the machine cycles for the majority of amplification reactions, since template build up prevents amplification in later cycles (Lu et al. 1995). It was in order to reduce the level of artefactual variation in my experiments that I used the EXPAND™ high fidelity PCR system instead of *Taq* polymerase for all of my amplification reactions (Sections 2.6.1 and 2.6.2). The EXPAND system contains two polymerase enzymes, *Taq* polymerase and *Pwo* polymerase, the latter possessing a 3’-5’ exonuclease proofreading activity. Consequently the fidelity of the EXPAND system (8.5×10^{-6} errors per base pair per PCR cycle) is three times higher than that of *Taq* polymerase. Although it is possible to achieve even lower error rates using a proofreading enzyme alone (Smith et al. 1997; Andre et al. 1997; Bracho et al. 1998) the PCR product yields are significantly lower. By using EXPAND instead of *Taq* polymerase and

performing 35 amplification cycles (20 effective cycles), I was able to obtain sufficient PCR product for the cloning reaction whilst minimising the number of polymerase misincorporations. There was no need to consider the PCR cycles performed post-cloning. Even if an error occurred in the first cycle of the colony PCR, only 50% of PCR products would contain a *Taq* polymerase-induced mutation. Using the PCR conditions described in Sections 2.6.1 and 2.6.2, 2% of 5'NCR PCR products and 4% of NS5b PCR products would theoretically contain a mutation due to polymerase infidelity. This equates to less than one band shift per 40 5'NCR clones and between 1 and 2 band shifts per 40 NS5b clones. This “background” level of variation is below that seen among clones from the majority of the individuals studied, indicating that most of the individuals were indeed carrying HCV as a quasispecies. I did not subtract this theoretical “background variation rate” from the actual number of variants detected by DGGE as I was satisfied that the observed number of variants was large and background error constant, for all of the studied individuals.

Although PCR is the main source of artefactual mutation, there are other sources. The enzyme used for the reverse transcription of the viral RNA (See Section 2.1.2) has an error rate of less than 1 in 28,000 nucleotides (Bracho et al. 1998). Since cDNA synthesis does not involve multiple cycles like PCR, this step does not contribute significantly to artefactual errors. Nucleotide sequencing itself is also prone to misincorporate bases owing to the requirement for a polymerase enzyme in the reaction. However this would not affect the number of observed variants in DGGE analysis but could lead to additional mutations in variants detected by DGGE and selected for sequence analysis.

In summary, DGGE is a fast and reliable method for assessing the quasispecies distribution of HCV infecting individuals. HCV quasispecies complexity and diversity varies between individuals, and the prevalence of mixed HCV infection appears to be low.

Chapter 5

HCV Quasispecies Evolution During Seroconversion

5.1 Introduction

The hypervariable region 1 (HVR-1) is a 27 amino acid segment at the amino terminus of the E2 gene (Figure 1.1) and it exhibits the highest sequence diversity of any region of the genome. Antibodies are directed against the region of the polyprotein encoded by HVR-1 (Weiner et al. 1992; Kato et al. 1993; Shimizu et al. 1994; Zibert et al. 1995; 1997). It is thought that the immune response targeted at this site drives the evolution of HVR-1, giving rise to the wide spectrum of HVR-1 variants that may be found in a given chronically infected individual (Weiner et al. 1992; van Doorn et al. 1995). This notion is supported by several reports of individuals with impaired immune function having reduced HVR-1 diversity in comparison to individuals with fully functioning immune systems (Odeberg et al. 1997; Toyoda et al. 1997; Lawal et al. 1997; Booth et al. 1998b). This is an observation that I have also made in the course of studying HVR-1 quasispecies distributions in the blood of chronically infected patients. Figure 5.1 shows HVR-1 sequence diversity, revealed by DGGE analysis, for: (a) a chronically infected injecting drug user, presumably immunocompetent, displaying a diverse quasispecies distribution; and (b) a patient who was immunosuppressed following a liver transplant, in whom only one dominant variant was detected.

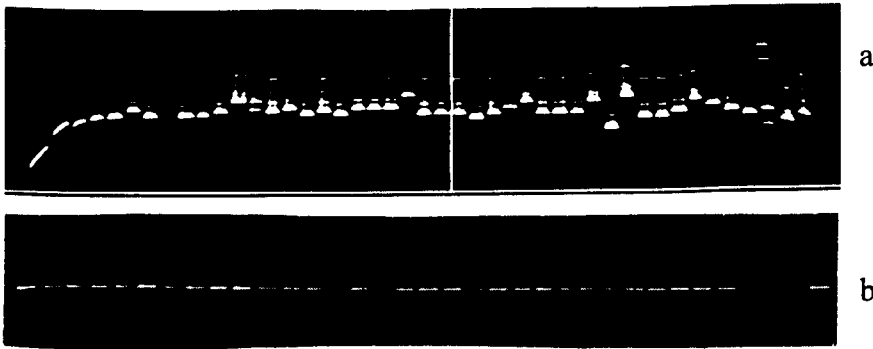


Figure 5.1: HVR-1 quasispecies distribution in (a) a chronically infected individual with liver disfunction and (b) a post-liver transplant patient.

It is not clear whether antibodies directed against the HVR-1 encoded peptide are sufficient to neutralise the virus (Kato et al. 1994; Shimizu et al. 1994; Zibert et al. 1997). A recent study suggests that the epitope expressed by this region acts as an immunological “decoy”, eliciting a strong immune response that is ineffective for viral clearance (Ray et al. 1999). The authors of this report demonstrated a correlation between clearance of virus and non-synonymous nucleotide change for the E1 region, implying that the immune response is both driving evolution and neutralising the virus. However, for the HVR-1, they showed a correlation between non-synonymous change in this region and persistence, not clearance. This led them to conclude that although the antibody response was driving HVR-1 evolution it was not sufficient to prevent chronic infection. Interestingly, an earlier study (Bassett et al. 1999) in chimpanzees showed that animals that cleared HCV lacked an E2 antibody response, whereas animals that became chronically infected had made antibodies to this region. If HVR-1 encoded peptides are indeed immunological decoys, stimulating an immune response at the expense of immune responses to other parts of the genome, then a lack of E2 antibody response could favour viral clearance.

The target of the host's immune response is the whole quasispecies population. Individual viral genomes evolve under the influence of the mutant spectrum of which they are part (Domingo et al. 1995). Variants that escape the viral swarm will evolve in a different environment that may favour the replication of hitherto minor components of the quasispecies (Domingo et al. 1995; Pawlotsky, 1998a).

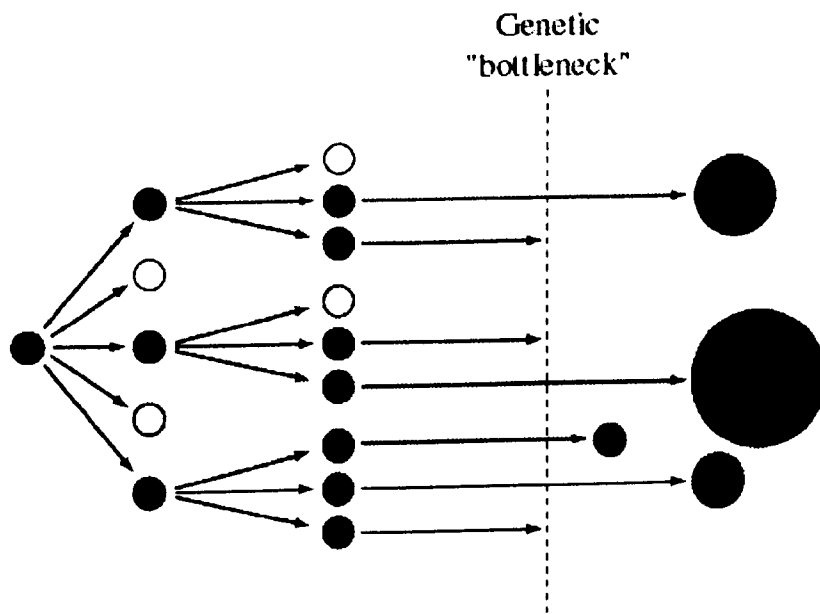


Figure 5.2: When a viral quasispecies faces a genetic bottle-neck only a few components will survive.

A genetic bottleneck describes a dramatic reduction in the size of a quasispecies population (Domingo et al. 1995; Pawlotsky, 1998a; Bergstrom et al. 1999). A transmission event is an example of a genetic bottleneck. Figure 5.2 is a schematic representation of what happens when a viral quasispecies passes through a genetic bottleneck. Only very large inoculum sizes will carry the full quasispecies distribution found in the transmitter. The smaller the inoculum size the tighter the bottleneck (Domingo et al. 1995; Pawlotsky, 1998a; Bergstrom et al. 1999). In addition to this, not all variants present in the inoculum will successfully invade target cells in the new host or replicate efficiently in a different

environment. The population introduced into the new host will comprise one or more “founder” strains which will establish a distinct quasispecies. A single variant will usually become dominant, although this may not be the same strain that prevailed in the previous host.

The concept of HCV transmission through a bottleneck has given rise to the consideration that quasispecies diversity during acute infection may be low. This is supported by two recent studies (Manzin et al. 1998; Ray et al. 1999), which showed the presence of very few viral variants carried by the recipient at the initial stages of HCV infection.

Generally, following transmission, mutations will begin to accumulate as a result of the neutral genetic drift of the virus due to error prone replication of its genome (Steinhauer and Holland, 1987). Synonymous nucleotide changes do not alter the fitness of a particular genome nor the quasispecies equilibrium. Non-synonymous nucleotide changes can be neutral too, but they can also be selected for. Negative selection occurs when amino acid mutations reduce the fitness of a variant. Either replication may be less efficient or the ability to escape immune clearance may be lost. These changes to the quasispecies are usually transient. Positive selection occurs when amino acid mutations increase the fitness of a variant. A minor variant may become dominant if it can replicate more efficiently than other variants or if it avoids immune clearance better (Domingo et al. 1995; Manzin et al. 1998; Ray et al. 1999). It is helpful to visualise a quasispecies population as existing in a “fitness landscape”. Mutations that increase fitness allow variants to move uphill to nearby fitness peaks, away from valleys where less fit variants exist (Wright, 1931; Eigen, 1992; Domingo et al. 1996).

This study was attempted to gain better insight into the diversity of HCV in early infection and how it evolves when the host begins to mount immunological attacks against the virus. The study examines changes in HVR-1 quasispecies distributions circulating in the blood stream of three individuals during very early infection, prior to and during HCV seroconversion. The DGGE strategy described in the preceding chapter of this thesis was utilised.

5.2 Materials and Methods

5.2.1 Patients

The three individuals in this study were in the acute phase of HCV infection. Patient RZ was a 50-year-old female who became jaundiced on returning from India. Serum samples from this patient were referred to the Virus Reference Division at Central Public Health Laboratory for testing. Infection was by HCV genotype 1b. Patient 11, a 24 year old male infected with HCV genotype 1a and patient 14, a 37 year old male infected with HCV genotype 1b, were plasma donors from whom serial samples were purchased from Bioclinical Partners Inc. (Franklin, MA, USA), Catalogue Numbers HCV6211 and HCV6214 respectively. Tables 5.1 – 5.3 show the results of HCV antibody testing by a third generation enzyme immunosorbent assay (EIA 3.0) (Ortho diagnostics, Chiron Corporation, Emeryville, CA, USA) and a third generation recombinant immunoblot assay (RIBA 3.0) (Ortho diagnostics, Chiron Corporation) of serial samples from the three patients. Increases in EIA and RIBA reactivity were indicators of acute infection. The levels of circulating HCV RNA, determined by the branched DNA (bDNA) assay (Chiron

Corporation) and alanine aminotransferase (ALT) levels were known in two of the patients only.

Time point	Date sample taken	HCV EIA 3.0 S/co	HCV RIBA 3.0 profile C100/C33/C22/NS5/SOD	HCV bDNA Copies/ml	ALT u/L (NR: 10 – 45)
1	18.01.99	2.18	Neg / 1+ / 2+ / neg / neg	Not done	Not done
2	22.02.99	13.44	+/- / 3+ / 4+ / +/- / neg	Not done	Not done

Table 5.1: Clinical information for serial samples taken from patient RZ. EIA readings are expressed as a ratio of the sample and the cut-off value for that run (S/co). NR= normal range.

Time point	Date sample taken	HCV EIA 3.0 (S/co)	HCV RIBA 3.0 profile C100/C33/C22/NS5/SOD	HCV bDNA (Copies/ml)	ALT u/L (NR: 10 – 45)
1	13.02.95	0.01	neg/ neg / neg/ neg / neg	21, 050,000	Normal
2	27.02.95	0.01	neg/ neg / neg / neg / neg	99,310,000	Normal
3	13.03.95	0.01	neg / neg / neg / neg / neg	>120,000,000	Normal
4	27.03.95	0.850	neg / 2+ / neg / neg / neg	>120,000,000	85
5	31.03.95	3.970	1+ / 3+ / neg / neg / neg	115,500,000	132
6	03.04.95	4.150	3+ / 4+ / neg / neg / neg	>120,000,000	Normal

Table 5.2: Clinical information for serial samples taken from patient 11. EIA readings are expressed as a ratio of the sample and the cut-off value for that run (S/co). NR= normal range.

Time point	Date sample taken	HCV EIA 3.0 (S/co)	HCV RIBA 3.0 profile C100/C33/C22/NS5/SOD	HCV bDNA (Copies/ml)	ALT u/L (NR: 10 – 45)
1	13.01.96	0.003	neg/ neg / neg/ neg / neg	6,357,000	Normal
2	23.01.96	0.003	neg/ neg / neg / neg / neg	6,910,000	Normal
3	05.02.96	0.005	neg / +/- / neg / neg / neg	5,374,000	Normal
4	14.02.96	2.643	+/- / 3+ / neg / neg / neg	6,278,000	Normal
5	02.03.96	4.126	2+ / 4+ / neg / neg / neg	2,446,000	116
6	09.03.96	4.126	2+ / 4+ / neg / neg / neg	2,031,000	152

Table 5.3: Clinical information for serial samples taken from patient 14. EIA readings are expressed as a ratio of the sample and the cut-off value for that run (S/co). NR= normal range.

5.2.2 Methods

A 189-base pair fragment from the E1 and E2 genes, encompassing HVR-1, was amplified from patient sera as described in Section 2.6.3. Following cloning (2.6.5 and 2.6.8), colony PCR products were analysed by DGGE as described in Sections 2.6.10 - 2.6.12. The nucleotide sequence of each variant was determined, as described in Section 2.6.13. Alignments were performed for both nucleotide and amino-acid sequences and ds/dn ratios were calculated at each time point for each of the three patients (2.6.14).

5.3 Results

5.3.1 HVR-1 quasispecies during acute infection

DGGE analysis allowed the changes in HVR-1 sequences over a short period of time in the three seroconverting individuals to be evaluated rapidly. Figures 5.3 - 5.22 show the quasispecies distributions for each patient at each time point. Overall, HVR-1 sequence diversity was low in these three patients, in comparison to what may be observed in a chronically infected individual (Figure 5.1(a)).

5.3.2 Patient RZ

The HVR-1 sequence diversity decreased in the later samples of this patient (Figures 5.3 and 5.4). Variant II, which accounted for eight out of thirty nine clonotypes in the earliest sample (Figure 5.3), was only represented three times among the thirty nine clones analysed one month later (Figure 5.4). Two unique variants, III and IV, seen at time point 1 (Figure 5.3) disappeared at time point 2, during which another unique variant, V, appeared (Figure 5.4).

Figures 5.5 and 5.6 show the nucleotide and amino-acid sequences of each of the variants that infected patient RZ. All variants differed from the dominant strain by at least one non-synonymous nucleotide change, although in three cases (IV, V and VI) this was just outside HVR-1. Table 5.4 shows the changes in ds/dn values at the two time points.

Time-point	ds value	dn value	ds/dn ratio
1	0.048	0.019	2.55
2	0.057	0.019	3.03

Table 5.4: Proportion of synonymous (ds) and proportion of non-synonymous (dn) changes among variants infecting Patient RZ at time-points one and two. Values were calculated as described in Section 2.6.14.

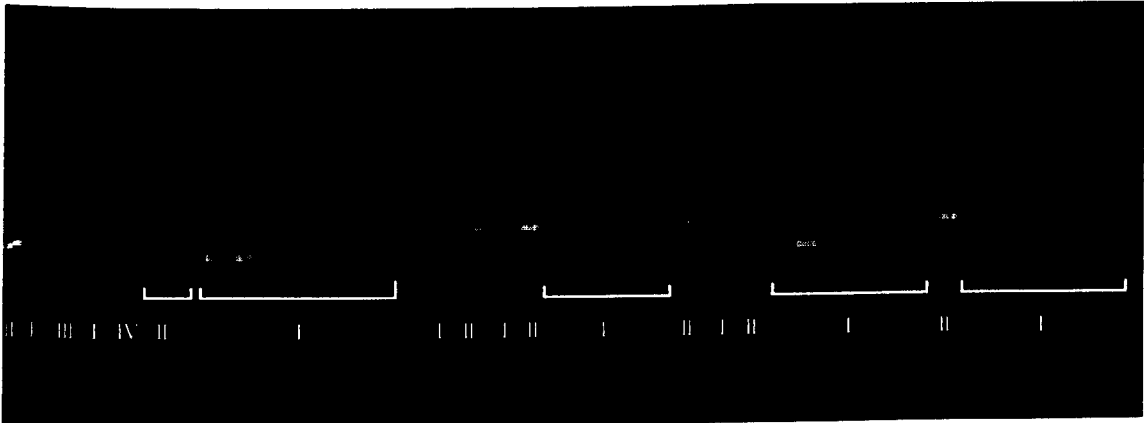


Figure 5.3: DGGE analysis of envelope region/HVR-1 quasispecies infecting Patient RZ at time point 1. Each clonotype is indicated by a different Roman numeral.

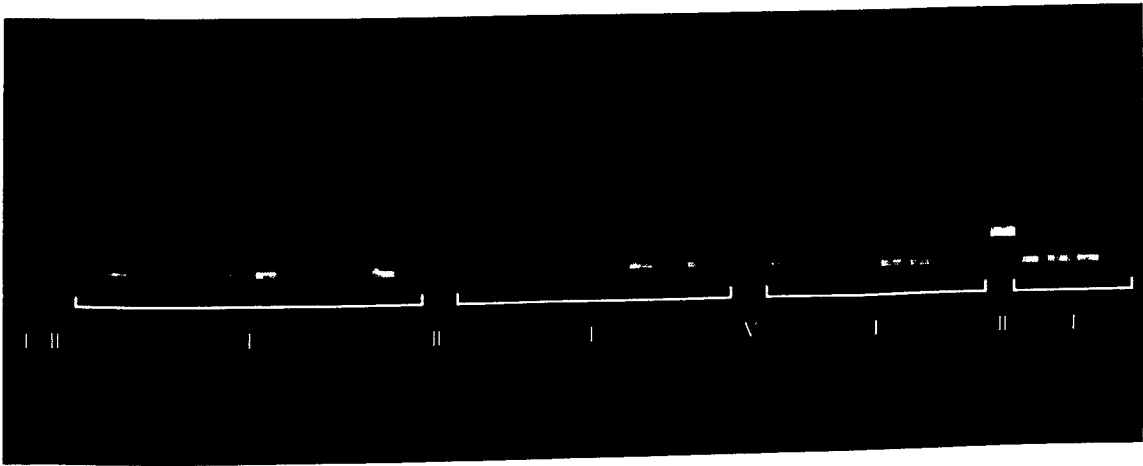


Figure 5.4: DGGE analysis of envelope region/HVR1 quasispecies infecting Patient RZ at time point 2. Each clonotype is indicated by a different Roman numeral.

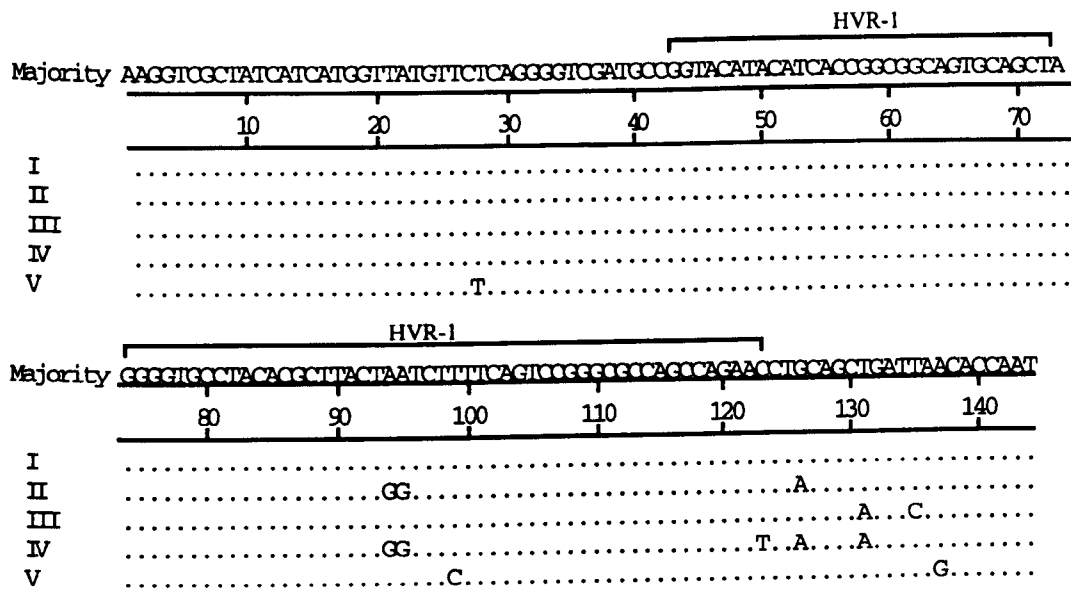


Figure 5.5: Nucleotide sequences of the envelope region/HVR1 clonotypes infecting Patient RZ. Nucleotides that match the consensus are indicated by ‘.’. Roman numerals match those in Figures 5.2 and 5.3. The brackets show the 81 base-pair HVR-1.

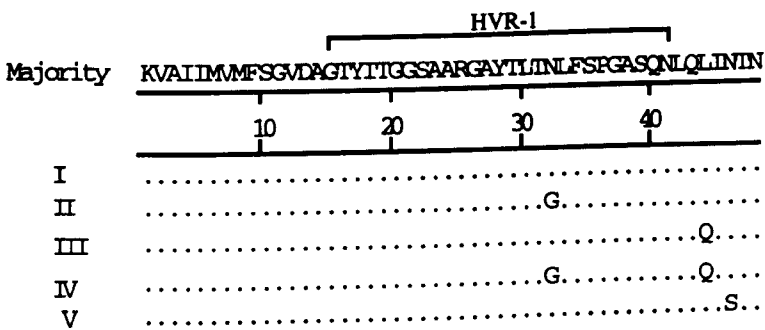


Figure 5.6: Amino acid sequences of the envelope region/HVR1 clonotypes infecting Patient RZ. Residues that match the consensus are indicated by ‘.’. Roman numerals match those in Figures 5.2 and 5.3. The brackets show the 27 amino acid HVR-1.

5.3.3 Patient 11

The first four samples from this patient were taken at two week intervals; the remaining samples were taken at four-day intervals (See Table 5.2). The DGGE profiles were very homogeneous in the first two samples. No variants were seen at time point 1 (Figure 5.7) and two weeks later only one of the thirty-four clonotypes was a variant (Figure 5.8). At time point 3 there were four variants out of thirty-eight clonotypes (Figure 5.9), but at time point 4 the distribution became homogeneous again, only one of the thirty-six clonotypes being a variant (Figure 5.10). At time point 5, again only one variant was seen (Figure 5.11). However, at the last time point, five of the forty clonotypes were variants (Figure 5.12). Interestingly, all of the variants seen in this patient occurred only once, with the exception of variant X, which was observed twice at time point 6.

Figures 5.13 and 5.14 show the nucleotide and amino-acid sequences of each of the variants found in Patient 11. Two variants (V and XI) had large deletions in this region of the genome and two others (III and VII) had a single-base deletion in their nucleotide sequences. The predicted protein sequences indicated that these variants were truncated, either as a result of a large deletion or the introduction of a stop codon. The amino-acid sequence of variant V was completely different, a result of the shift in the reading frame (Figure 5.14). All but three of the variants (VIII, IX and XII) differed from the dominant species by at least one non-synonymous nucleotide change, although in two cases (II and VI) this occurred just outside HVR-1 (Figures 5.13 and 5.14). Table 5.5 shows the changes in ds/dn values at the six time points.

Time-point	ds value	dn value	ds/dn ratio
1	0.000	0.000	0.00
2	0.000	0.010	0.00
3	0.000	0.015	0.00
4	0.000	0.010	0.00
5	0.026	0.000	∞
6	0.040	0.005	8.31

Table 5.5: Proportion of synonymous (ds) and proportion of non-synonymous (dn) changes among variants infecting Patient 11 at time-points 1 to 6. Values were calculated as described in Section 2.6.14.

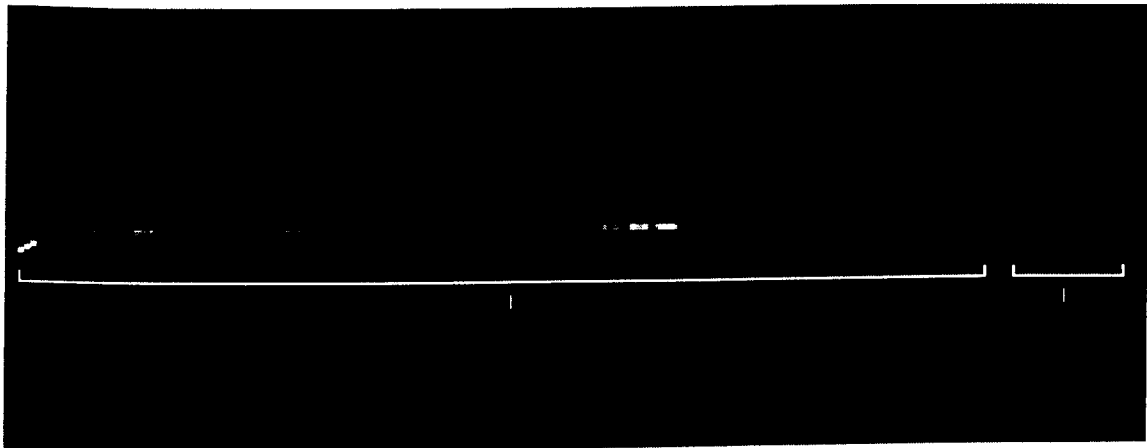


Figure 5.7: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 1. Each clonotype is indicated by a different Roman numeral.

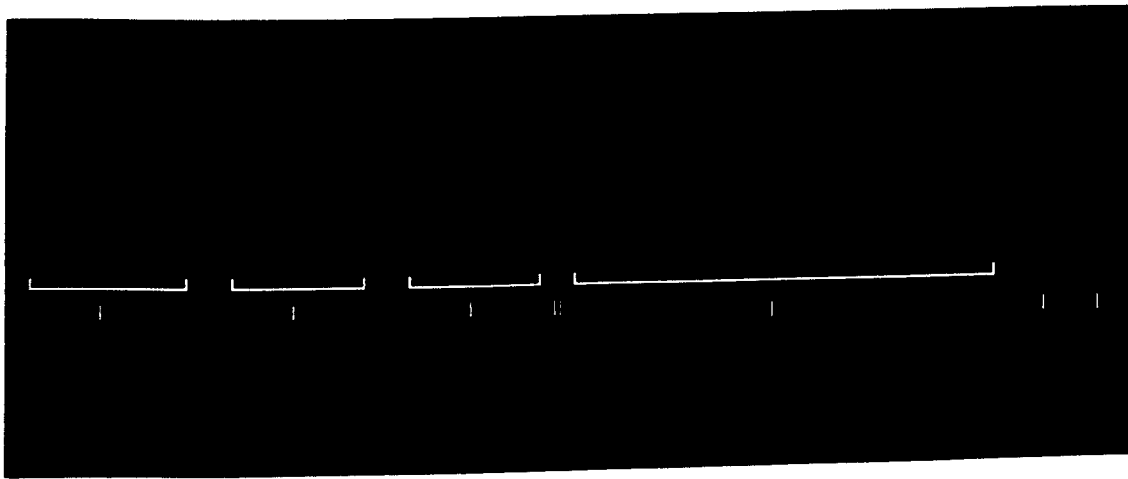


Figure 5.8: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 2. Each clonotype is indicated by a different Roman numeral.

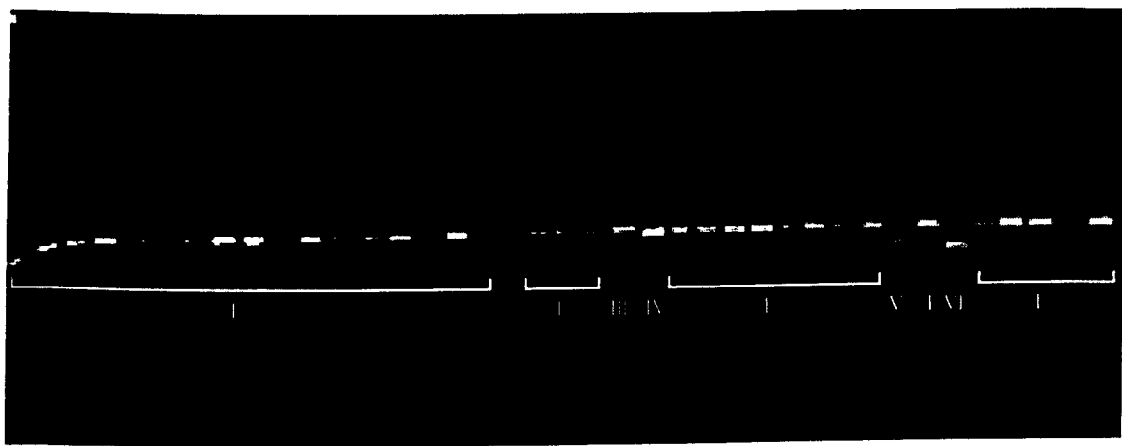


Figure 5.9: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 3. Each clonotype is indicated by a different Roman numeral.



Figure 5.10: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 4. Each clonotype is indicated by a different Roman numeral. The lane marked with '*' does not contain a colony PCR product derived from the HVR1, as determined by sequencing.

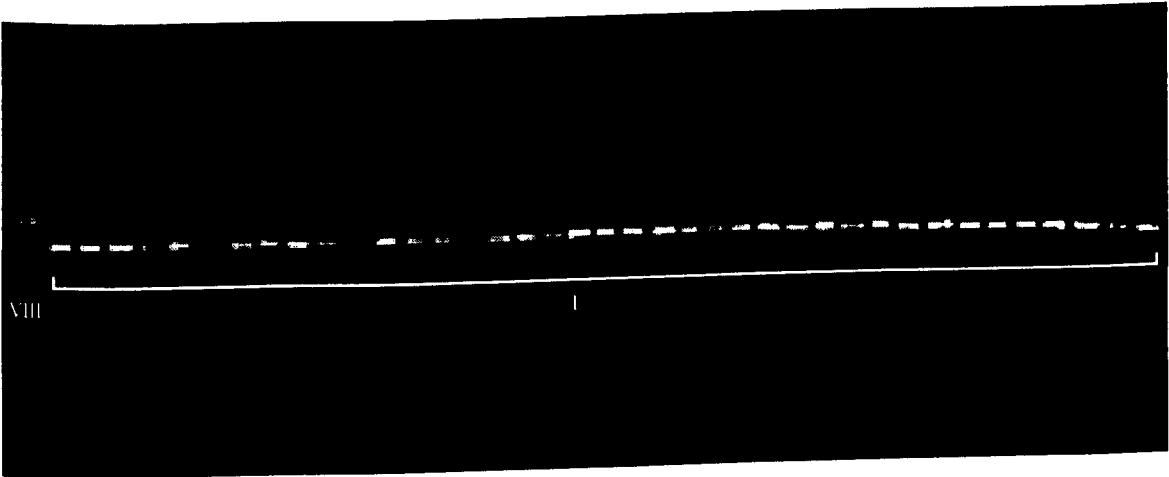


Figure 5.11: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 5. Each clonotype is indicated by a different Roman numeral.

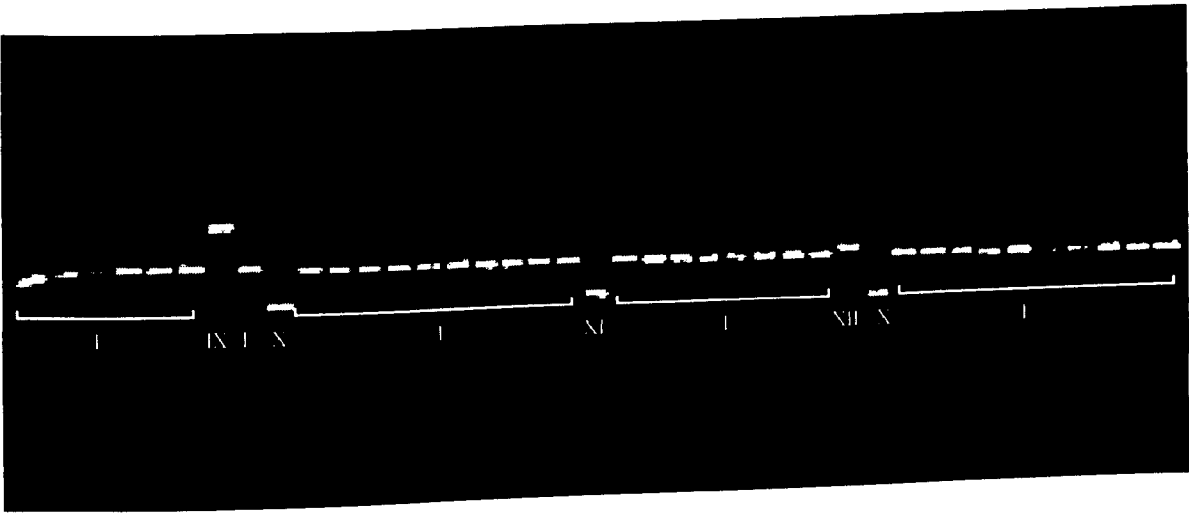


Figure 5.12: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 11 at time point 6. Each clonotype is indicated by a different Roman numeral.

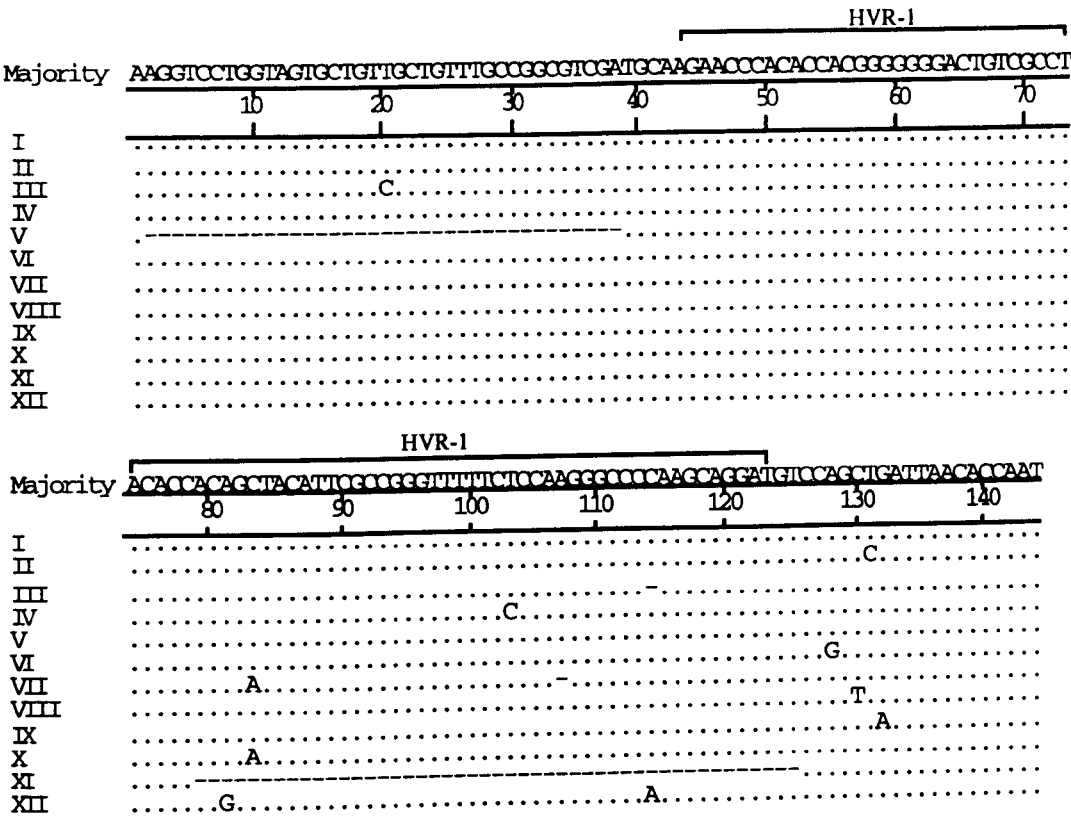


Figure 5.13: Nucleotide sequences of the envelope/HVR1 clonotypes infecting Patient 11. Nucleotides that match the consensus are indicated by ‘.’. Roman numerals match those in Figures 5.6 - 5.11. The brackets show the 81 base-pair HVR-1.

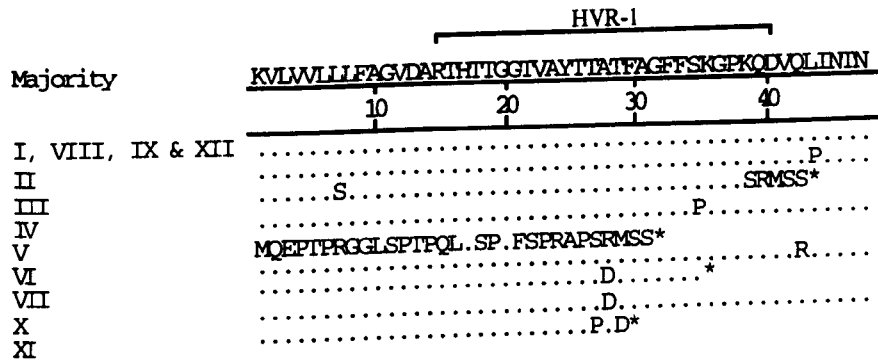


Figure 5.14: Amino acid sequences of the envelope/HVR1 clonotypes infecting Patient 11. Residues that match the consensus are indicated by ‘.’. Stop codons are represented as ‘*’. Roman numerals match those in Figures 5.6 - 5.11. The brackets show the 27 amino acid HVR-1.

5.3.4 Patient 14

This patient appeared to be infected with a greater variety of HCV variants than the other two patients in this study. There was an overall increase in the number of variants seen during the course of seroconversion (Figures 5.15 - 5.20).

Serum samples were taken from Patient 14 at intervals of 7-14 days (See Table 5.3). Specimens from time points 1 and 2 displayed the most homogenous distributions (Figures 5.15 and 5.16), with two out of forty and one out of thirty-nine being variant, respectively. At each of the time points 3 and 4, five out of thirty-nine clonotypes were variants (Figures 5.17 and 5.18). At time point 5 only two out of thirty-nine clonotypes were variants (Figure 5.19) but at time point 6 the number of variants had increased to seven out of thirty-eight (Figure 5.20).

Most of the minor variants infecting Patient 14 were unique, but variants II and XIII could be recognised at several time points. Variant II occurred at least once at time points 1, 3, 5 and 6 (Figures 5.15, 5.17, 5.19 and 5.20). Variant XIII, which emerged at time point 4, was seen at time points 5 and 6 (Figures 5.18 - 5.20).

Figures 5.21 and 5.22 show the nucleotide and amino-acid sequences of each of the variants carried by Patient 14. Many variants shared sequence motifs and it appeared that at least two groups of variants existed in this patient. Phylogenetic analysis was performed using the CLUSTAL V algorithm of the MEGALIGN program (Section 2.6.14). The dendrogram in Figure 5.23 shows that the variants that infected Patient 14 can be classified as three groups. Variants I, IV, V, VI, VII, X, XI, XII, XV and XVI segregated to one group, called

group A here. Variants XIV and XVII also clustered with this group, although they are less closely related. Variants VIII, IX and XIII segregated to group B. Group C comprised variants II and III.

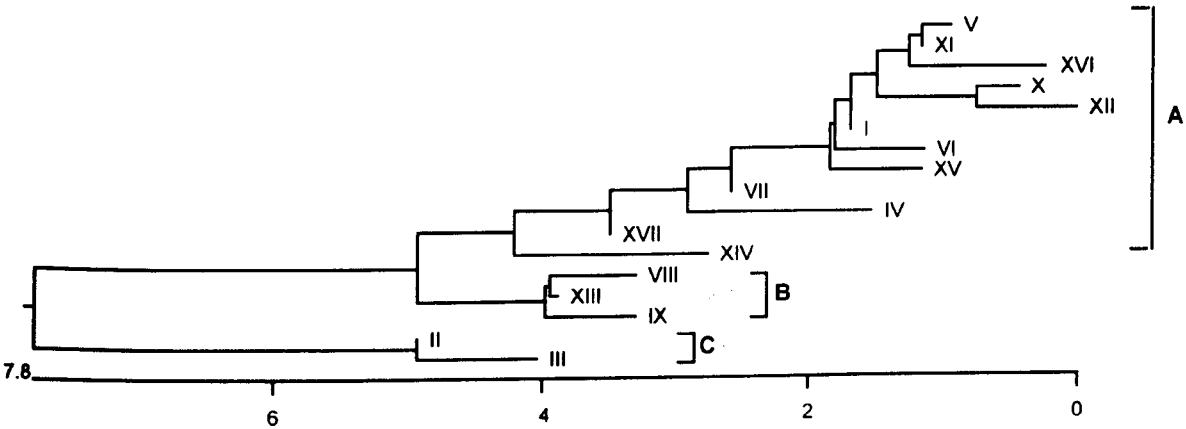


Figure 5.23: Dendrogram to show relatedness of HVR-1 variants infecting Patient 14. The variants cluster into three groups, A, B and C. The tree was produced using the CLUSTAL V algorithm of the MEGALIGN program.

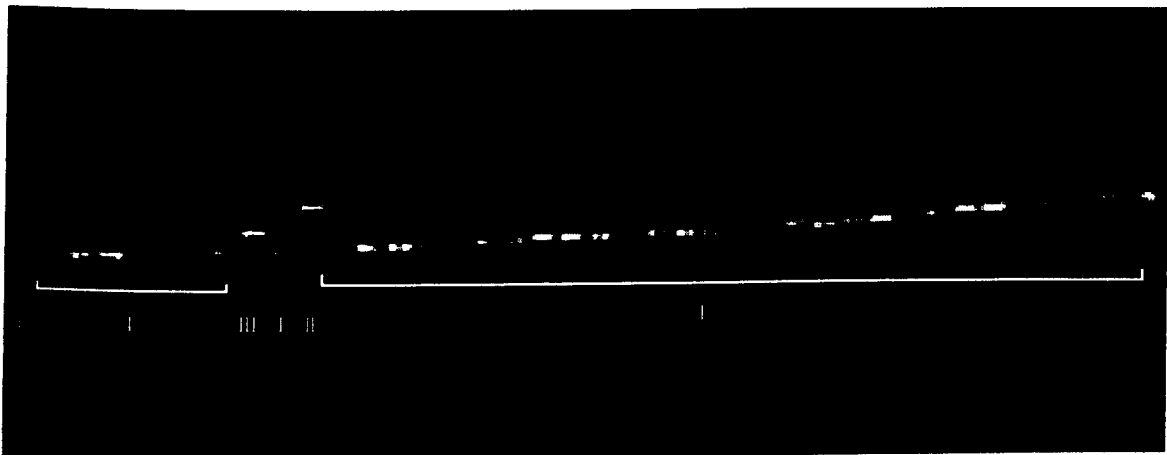


Figure 5.15: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 14 at time point 1. Each clonotype is indicated by a different Roman numeral.



Figure 5.16: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 14 at time point 2. Each clonotype is indicated by a different Roman numeral.

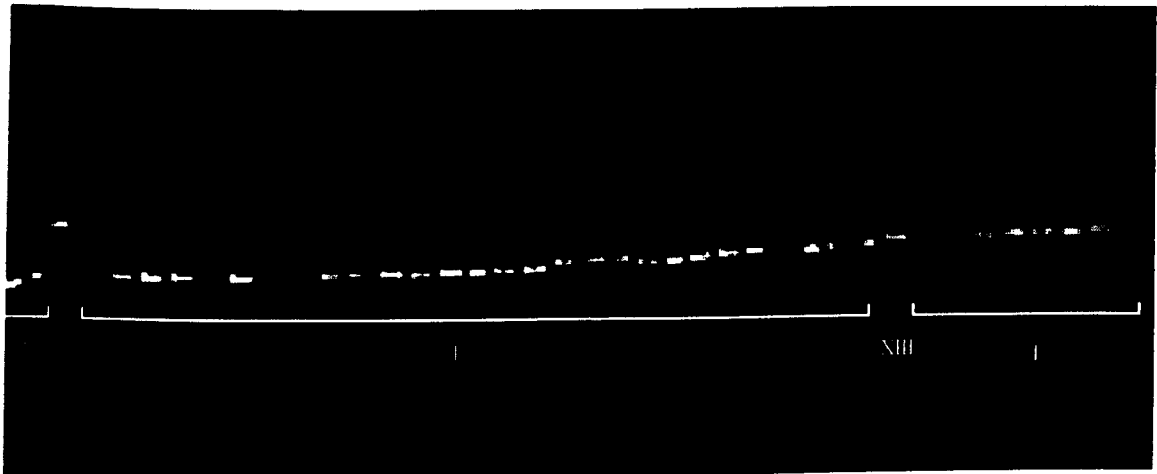


Figure 5.19: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 14 at time point 5. Each clonotype is indicated by a different Roman numeral.



Figure 5.20: DGGE analysis of envelope/HVR1 quasispecies infecting Patient 14 at time point 6. Each clonotype is indicated by a different Roman numeral.

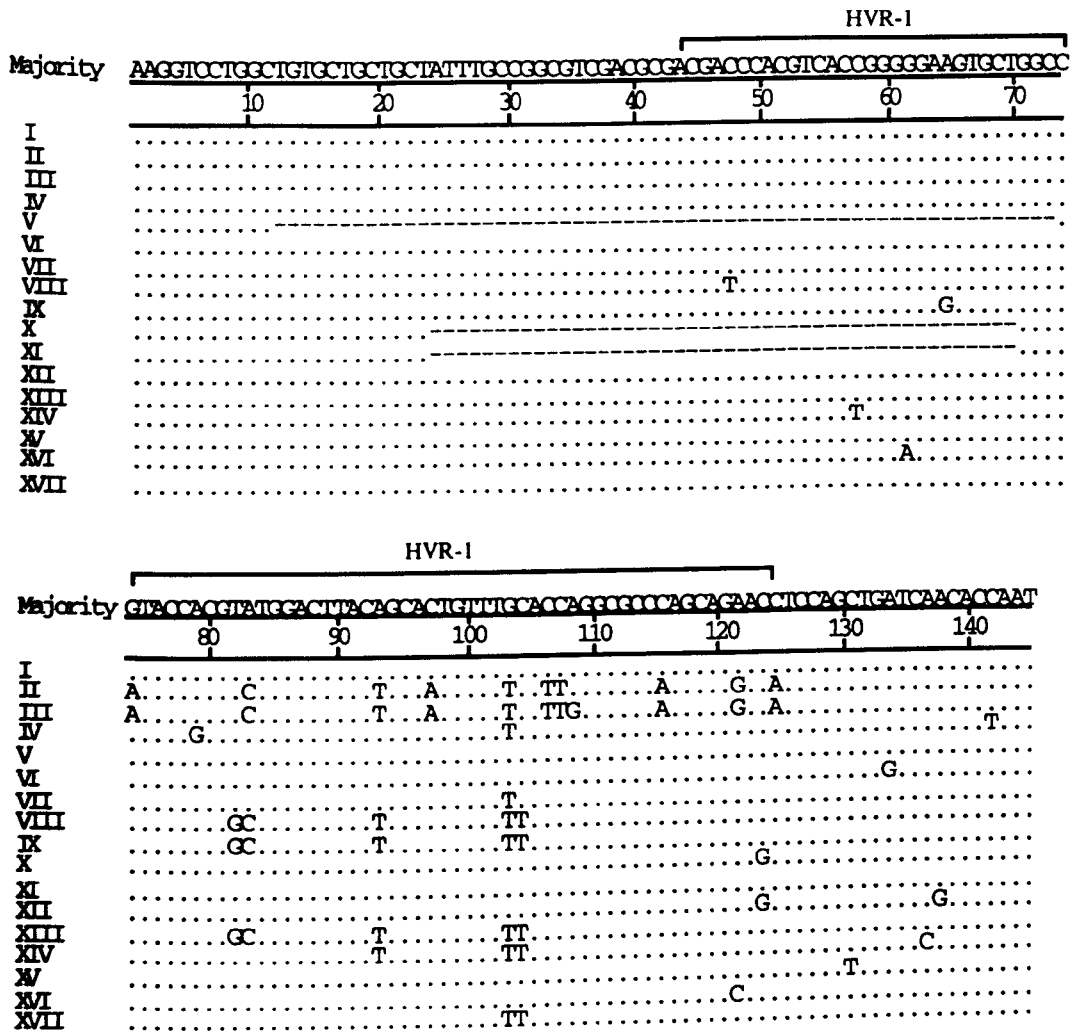


Figure 5.21: Nucleotide sequences of the envelope/HVR1 clonotypes infecting Patient 14. Nucleotides that match the consensus are indicated by '.'. Roman numerals match those in Figures 5.14 - 5.19. The brackets show the 81 base-pair HVR-1.

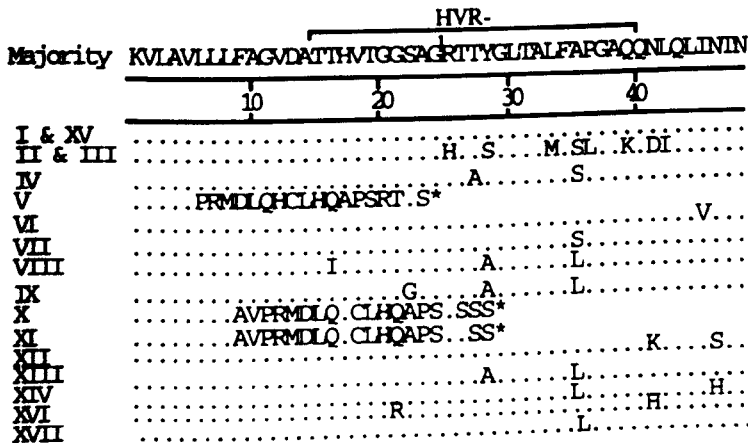


Figure 5.22: Amino acid sequences of the envelope/HVR-1 clonotypes infecting Patient 14. Residues that match the consensus are indicated by '.'. Stop codons are represented as '*'. Roman numerals match those in Figures 5.2 and 5.3. The brackets show the 27 amino acid HVR-1.

Three variants (V, X and XI) had large deletions in this region of the genome. This resulted in both truncation of the predicted protein and widely divergent amino acid sequences due to the shift in the reading frame (Figure 5.22). All but one (XV) of the variants differed from the dominant strain by at least one non-synonymous nucleotide change, although in one case (VI) this change occurred just outside HVR-1 (Figures 5.21 and 5.22). Table 5.6 shows the changes in ds/dn values at the six time points.

Time-point	ds value	dn value	ds/dn ratio
1	0.043	0.058	0.74
2	0.024	0.020	1.23
3	0.020	0.055	0.36
4	0.016	0.042	0.39
5	0.025	0.071	0.35
6	0.032	0.046	0.70

Table 5.6: Proportion of synonymous (ds) and proportion of non-synonymous (dn) changes among variants infecting patient 14 at time-points one to six. Values were calculated as described in Section 2.6.14.

5.4 Discussion

This study of HCV quasispecies during acute infection, investigated by DGGE analysis of a substantial number of clones derived from the HVR-1, shows that quasispecies diversity immediately following transmission does vary between individuals and that sequence changes in HVR-1 during acute infection may be under different types of selective pressures.

5.4.1 Variation in HCV diversity in the early phase of acute infection

Patients RZ and 11 displayed relatively homogeneous quasispecies distributions during early HCV infection. This picture, similar to what was observed in two other studies of

quasispecies diversity during acute HCV infection (Manzin et al. 1998), is consistent with founder strains having passed through a genetic bottleneck during transmission and evolving in the new host to establish distinct viral quasispecies.

In contrast, Patient 14 was infected with several founder strains that may be grouped into three clades, indicated as A, B and C in the dendrogram shown in Figure 5.23. It is possible that this individual was exposed to a large volume of infectious material and hence a fuller representation of the quasispecies circulating within the transmitter. Large inoculum sizes are known to increase the likelihood of variants capable of establishing infection after being transmitted to the new host (Bergstrom et al. 1999).

The observations in Patient 14 have implications for the management of acute HCV infection. If HCV transmission is an oligoclonal event, then there appears to be more opportunity for early intervention with antiviral drugs to prevent infection or progression to chronic infection (Manzin et al. 1998). However, the diversity of HCV in the inoculum may not always be so restricted, as shown by Patient 14. Whether patients (such as patient 14) who are initially infected with a wider variety of HCV types are more refractory to antiviral treatment requires further research.

5.4.2 Variation in HCV evolution during the course of acute infection

The virus infecting Patient RZ appears to be under negative selective pressure, since the ds/dn values were larger than one at both time points (Table 5.4). For Patient 11, ds/dn values in the first four samples were all less than one (Table 5.5), suggesting that the viral quasispecies infecting this patient at this time was subjected to positive selective pressure.

At time points five and six the high ds/dn values imply that the HVR-1 was evolving under negative selective pressure. For Patient 14, ds/dn values for the first two time points indicate neutral evolution. The remaining four time points were associated with ds/dn values that were consistent with positive selective pressure being exerted (Table 5.6). However, the number of variants observed at each time point for all three patients was small. Coupled with the relatively small size of the amplified fragment this makes the total number of nucleotides analysed for each patient somewhat small. Consequently, the ds/dn values calculated above may not provide a very accurate model of viral evolution. Other studies, such as those by Manzin et al (1998) and Ray et al (1999), have not only included a larger number of individuals, they have also analysed a larger portion of the envelope gene over a longer period of time.

Figures 5.24 and 5.25 show how, for Patients 11 and 14, the ds/dn ratios varied with time in relation to antibody levels, ALT values and viral load. Viral load appears to correlate with ds/dn ratio for both patients. It is unlikely that an increase or decrease in viral replication would affect the proportion of synonymous and non-synonymous mutations. However, the most striking feature of these charts is how the ds/dn ratio correlated with other variables. Changes in ds/dn ratio did not correlate with changes in ALT, and more importantly did not correlate with antibody level changes. For Patient 11 antibody level increased as the ds/dn ratio increased whereas the reverse was true for Patient 14.

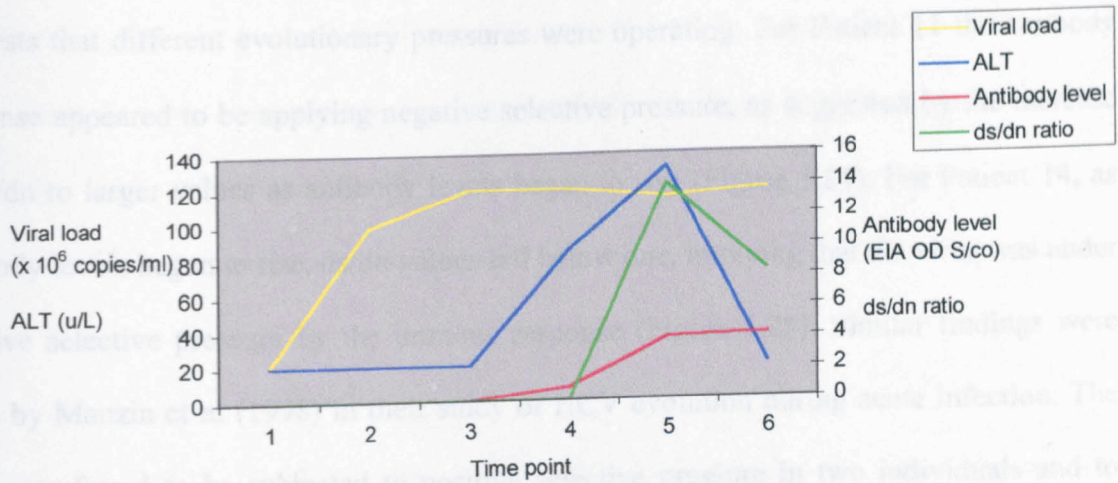


Figure 5.24: Chart to show how viral load, ALT values, antibody levels and ds/dn ratios change during acute infection in Patient 11.

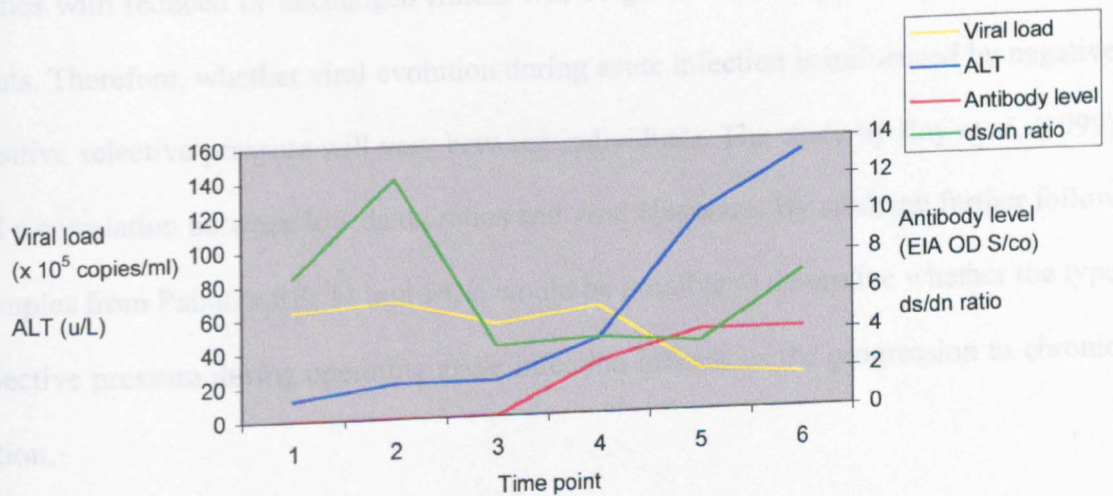


Figure 5.25: Chart to show how viral load, ALT values, antibody levels and ds/dn ratios change during acute infection in Patient 14.

The conflicting relationship between antibody level and ds/dn ratio in these two patients suggests that different evolutionary pressures were operating. For Patient 11 the antibody response appeared to be applying negative selective pressure, as suggested by the increase in ds/dn to larger values as antibody levels began to rise (Figure 5.24). For Patient 14, as antibody levels began to rise, ds/dn values fell below one, implying that the virus was under positive selective pressure by the immune response (Figure 5.25). Similar findings were made by Manzin et al (1998) in their study of HCV evolution during acute infection. The virus was found to be subjected to positive selective pressure in two individuals and to negative selective pressure in another. In aggregate, these results suggest that the evolution of HCV during acute infection differs between individuals and do not support the general hypothesis that the immune response drives the evolution of the virus to produce antibody escape mutants (Kato et al. 1993; Shimizu et al. 1994). Since mutations accumulate randomly in the HCV genome over time, the result of error prone replication, many genomes with reduced or unchanged fitness will be generated along with antibody escape mutants. Therefore, whether viral evolution during acute infection is influenced by negative or positive selective pressure will vary between individuals. The study by Ray et al. (1999) found a correlation between low ds/dn ratios and viral clearance. By studying further follow up samples from Patients RZ, 11 and 14, it would be possible to determine whether the type of selective pressure during operating acute infection influences the progression to chronic infection.

5.4.3 Limitations of experimental methods

Errors introduced during PCR can lead to an overestimation of quasispecies diversity. The problem of Taq polymerase misincorporation was discussed in Chapter 4 and the same

considerations apply here. Using the error rate for EXPAND polymerase described in Section 4.3.5, the 45 cycles of PCR used to amplify the 200 base pair HVR-1 amplicon (See section 2.6) would be expected to generate just over 2% of PCR products with a *Taq* polymerase misincorporation. As the proportion of minor variants seen is more than 5% for each of the three patients, it is unlikely that the variation detected was due to misincorporation errors.

Although the DGGE method is able to screen a large number of clones from each specimen, this number is still finite. Therefore, DGGE will not identify all minor variants present in a sample: there will be other minor strains that were transmitted to these three patients that cannot be identified. Similarly, not all variants arising through viral evolution will be detected.

Chapter 6

Summary and Future Studies

6.1 Summary of thesis

6.1.1 Overview of experimental design

The genetic diversity and evolution of HCV has been investigated. Chapter 3 described the distribution of HCV genotypes in England and Wales. A modified PCR-RFLP procedure allowed a large number of samples to be processed rapidly. The survey identified the prevalence of the six HCV genotypes and various subtypes in samples from England and Wales. However, limited information about the infected individuals meant that it was not possible to deduce whether infections by genotypes that are uncommon in northern Europe were acquired in other countries.

Chapter 4 investigated the quasispecies variation of HCV infecting 15 individuals. DGGE analysis of cloned PCR products, followed by sequencing of variants, allowed 40 clones from two genomic regions from each individual to be studied. Although this method can identify a relatively broad spectrum of viral quasispecies more rapidly than sequencing alone, this is still only a fraction of the true number of variants present. Many minor variants exist that may never be detected. The quasispecies variation was correlated with patient group (haemophilia patients, blood donors or IDUs) and information about genotype and age group was also available. However, limited clinical information meant that no association between quasispecies diversity and markers liver disease or duration of infection could be made.

The evolution of HCV during acute infection was investigated in Chapter 5. Only three patients were included in the study. Individuals acutely infected with HCV were difficult to identify and the HVR-1 could only be amplified from a few such samples. Samples from more patients at a greater number of time points, and amplification of a larger fragment of the envelope gene, would have allowed a more accurate picture of HCV evolution during acute infection to have been drawn.

6.1.2 Conclusions

The thesis set out to test three hypotheses. 1: The distribution of HCV genotypes in England and Wales is similar to that found in the rest of northern Europe; 2: The mode of acquisition affects both the genotype and complexity of the circulating virus, multiply exposed individuals being more likely to be infected by more than one genotype; and 3: the genetic variation of HCV does not fluctuate greatly during the first few weeks of infection. The following conclusions were drawn.

6.1.2.1 Distribution of HCV genotypes in England and Wales

The majority of HCV infections in England and Wales are caused by genotypes 3a (37%), 1a (32%) and 1b (15%). The genotype distributions in individual patient groups are similar to the overall distribution, with the exception of haemophilia patients in whom type 3a is less prevalent. This pattern suggests a common route of transmission for HCV in sub-populations, other than haemophilia patients, within England and Wales. It is suggested that intravenous drug use is that common route. The high frequency of type 3a infections has implications for HCV screening in England and Wales, since serological tests used are sub-optimal for the detection of genotype 3 infections.

6.1.2.2 Infection with multiple HCV genotypes or subtypes is rare

Infection by more than one genotype or subtype was detected in only 1% of 567 individuals, using a PCR-RFLP genotyping technique known to detect one genome belonging to a minor HCV variant that is present in every fifty genomes belonging to the dominant variant. In addition to this, none of the 40 clones derived from the 5'NCR and NS5b subgenomic regions in each of the 15 individuals belonged to different genotypes or subtypes. Thus multiple genotypes and subtypes rarely co-circulate, although chronic infection with HCV does not protect against superinfection. It is possible that an existing viral population can suppress the replication of a newly introduced strain. An important consequence of this finding is that a live, attenuated HCV vaccine may be effective in preventing primary HCV infection, although it is likely that the risk of reversion to virulence would be too high for this to be a realistic strategy.

6.1.2.3 The direction of evolution of HCV during acute infection is variable

Transmission of HCV may occur through a genetic bottleneck, and for this reason it has been suggested that viral diversity during early infection is low. The findings described in Chapter 5 contradict this idea, as one of the three acutely infected patients studied harboured a relatively diverse quasispecies population. Studies of the evolution of HCV in acute infection therefore need to take into account the diversity already present in the inoculum.

It is hypothesised that HCV evolves to escape the host's immune response, particularly during acute infection, in order to persist. I found in the sequential sera of patients undergoing HCV seroconversion that there was evidence of both positive and negative

selection in operation, as well as neutral evolution. Therefore, the evolutionary pressures in acutely infected patients are not uniform.

6.2 Future studies

To determine the true prevalence of mixed genotype or subtype HCV infections a more sensitive method for genotyping is necessary. Type specific PCR followed by nucleotide sequencing could be employed to perform both cross-sectional and longitudinal analyses of genotypes in individuals at high risk of multiple exposure to HCV. Cloning/DGGE analysis of the whole envelope gene amplified from HCV infecting a large panel of seroconverters would be required to correlate the direction of HCV evolution with inoculum heterogeneity, antibody and CTL responses, changes in viraemia levels and the host's genetic make-up. Longer-term follow up studies should allow the outcome of acute infection (whether leading to clearance or chronic infection) to be predicted.

References

- Andre, P., Kim, A., Khrapko, K. and Thilly, W.G. (1997) Fidelity and mutational spectrum of Pfu DNA polymerase on a human mitochondrial DNA sequence. *PCR Methods & Applications* **7**, 843-852.
- Barrera, J.M., Francis, B., Ercilla, G., Nelles, M., Achord, D., Darner, J., Lee and SR (1995) Improved detection of anti-HCV in post-transfusion hepatitis by a third-generation ELISA. *Vox Sanguinis* **68**, 15-18.
- Barrett, A.D. and Dimmock, N.J. (1986) Defective interfering viruses and infections of animals. *Current Topics in Microbiology & Immunology* **128**, 55-84.
- Bartenschlager, R. (1997) Candidate targets for hepatitis C virus-specific antiviral therapy. *Intervirology* **40**, 378-393.
- Basaras, M., Lombera, N., de las, H., Lopez, C., Arrese, E. and Cisterna, R. (1997) Distribution of HCV genotypes in patients infected by different sources. *Research in Virology*. **148**, 367-373.
- Bassett, S.E., Brasky, K.M. and Lanford, R.E. (1998) Analysis of hepatitis C virus-inoculated chimpanzees reveals unexpected clinical profiles. *Journal of Virology*. **72**, 2589-2599.
- Bassett, S.E., Thomas, D.L., Brasky, K.M. and Lanford, R.E. (1999) Viral persistence, antibody to E1 and E2, and hypervariable region 1 sequence stability in hepatitis C virus-inoculated chimpanzees. *Journal of Virology* **73**, 1118-1126.
- Behrens, S.E., Tomei, L. and De Francesco, R. (1996) Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus. *EMBO Journal* **15**, 12-22.
- Berg, T., Hopf, U., Stark, K., Baumgarten, R., Lobeck, H. and Schreier, E. (1997) Distribution of hepatitis C virus genotypes in German patients with chronic hepatitis

- C: correlation with clinical and virological parameters. *Journal of Hepatology* **26**, 484-491.
- Bergstrom, C.T., McElhany, P. and Real, L.A. (1999) Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *PNAS* **96**, 5095-5100.
- Booth, J.C. (1998a) Chronic hepatitis C: the virus, its discovery and the natural history of the disease. *Journal of Viral Hepatitis* **5**, 213-222.
- Booth, J.C., Kumar, U., Webster, D., Monjardino, J. and Thomas, H.C. (1998b) Comparison of the rate of sequence variation in the hypervariable region of E2/NS1 region of hepatitis C virus in normal and hypogammaglobulinemic patients. *Hepatology* **27**, 223-227.
- Bracho, M.A., Moya, A. and Barrio, E. (1998) Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *Journal of General Virology* **79**, 2921-2928.
- Brackmann, S., Gerritzen, A. and Oldenburg, J. (1993) Search for intrafamilial transmission of hepatitis C virus in haemophilia patients. *Blood* **81**, 1077-1082.
- Brambilla, S., Bellati, G., Asti, M., Lisa, A., Candusso, M.E., D'Amico, M., Grassi, G., Giacca, M., Franchini, A., Bruno, S., Ideo, G., Mondelli, M.U. and Silini, E.M. (1998) Dynamics of hypervariable region 1 variation in hepatitis C virus infection and correlation with clinical and virological features of liver disease. *Hepatology* **27**, 1678-1686.
- Bruno, S., Silini, E., Crosignani, A., Borzio, F., Leandro, G., Bono, F., Asti, M., Rossi, S., Larghi, A., Cerino, A., Podda, M. and Mondelli, M.U. (1997) Hepatitis C virus genotypes and risk of hepatocellular carcinoma in cirrhosis: a prospective study. *Hepatology* **25**, 754-758.
- Bukh, J., Miller, R.H. and Purcell, R.H. (1995) Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Seminars in Liver Disease* **15**, 41-63.

- Bukh, J., Purcell, R.H. and Miller, R.H. (1992) Sequence analysis of the 5' noncoding region of hepatitis C virus. *PNAS* **89**, 4942-4946.
- Cabot, B., Esteban, J.I., Martell, M., Genesca, J., Vargas, V., Esteban, R., Guardia, J. and Gomez, J. (1997) Structure of replicating hepatitis C virus (HCV) quasispecies in the liver may not be reflected by analysis of circulating HCV virions. *Journal of Virology* **71**, 1732-1734.
- Chang, J., Yang, S.H., Cho, Y.G., Hwang, S.B., Hahn, Y.S. and Sung, Y.C. (1998) Hepatitis C virus core from two different genotypes has an oncogenic potential but is not sufficient for transforming primary rat embryo fibroblasts in cooperation with the H-ras oncogene. *Journal of Virology* **72**, 3060-3065.
- Chayama, K., Tsubota, A., Kobayashi, M., Okamoto, K., Hashimoto, M., Miyano, Y., Koike, H., Koida, I., Arase, Y., Saitoh, S., Suzuki, Y., Murashima, N., Ikeda, K. and Kumada, H. (1997) Pretreatment virus load and multiple amino acid substitutions in the interferon sensitivity-determining region predict the outcome of interferon treatment in patients with chronic genotype 1b hepatitis C virus infection. *Hepatology* **25**, 745-749.
- Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**, 359-362.
- Choo, Q.L., Richman, K.H., Han, J.H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina-Selby, R. and Barr, P.J. (1991) Genetic organization and diversity of the hepatitis C virus. *PNAS* **88**, 2451-2455.
- Chung, R.T., Monto, A., Dienstag, J.L. and Kaplan, L.M. (1999) Mutations in the NS5A region do not predict interferon-responsiveness in american patients infected with genotype 1b hepatitis C virus. *Journal of Medical Virology* **58**, 353-358.
- Clewley, J.P. (1998a) A user's guide to producing and interpreting tree diagrams in taxonomy and phylogenetics. Part 2. The multiple alignment of DNA and protein

- sequences to determine their relationships. *Communicable Disease & Public Health* **1**, 132-134.
- Clewley, J.P. (1998c) A user's guide to producing and interpreting tree diagrams in taxonomy and phylogenetics. Part I. Introduction and naming of parts. *Communicable Disease & Public Health* **1**, 64-66.
- Coffin, J.M. (1995) HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267**, 483-489.
- Cohen, J. (1999) The scientific challenge of hepatitis C. *Science* **285**, 26-30.
- Cotton, R.G. (1993) Current methods of mutation detection. *Mutation Research* **285**, 125-144.
- Courouce, A.M., Le Marrec, N., Girault, A., Ducamp, S. and Simon, N. (1994) Anti-hepatitis C virus (anti-HCV) seroconversion in patients undergoing hemodialysis: comparison of second- and third-generation anti-HCV assays. *Transfusion* **34**, 790-795.
- Cremonesi, L., Firpo, S., Ferrari, M., Righetti, P.G. and Gelfi, C. (1997) Double-gradient DGGE for optimized detection of DNA point mutations. *Biotechniques* **22**, 326-330.
- Dammacco, F., Gatti, P. and Sansonno, D. (1998) Hepatitis C virus infection, mixed cryoglobulinemia, and non-Hodgkin's lymphoma: an emerging picture. *Leukemia & Lymphoma* **31**, 463-476.
- Davidson, F., Simmonds, P., Ferguson, J.C., Jarvis, L.M., Dow, B.C., Follett, E.A., Seed, C.R., Krusius, T., Lin, C. and Medgyesi, G.A. (1995) Survey of major genotypes and subtypes of hepatitis C virus using RFLP of sequences amplified from the 5' non-coding region. *Journal of General Virology* **76**, 1197-1204.
- Davis, G.L., Beck, J.R., Farrell and Poynard, T. (1998) Prolonged treatment with interferon in patients with histologically mild chronic hepatitis C: a decision analysis. *Journal of Viral Hepatitis* **5**, 313-321.

- de la Torre, J.C. and Holland, J.J. (1990) *Journal of Virology* **64**, 6278-6281.
- Domingo, E. (1998a) Quasispecies and the implications for virus persistence and escape. *Clinical & Diagnostic Virology* **10**, 97-101.
- Domingo, E., Baranowski, E., Ruiz-Jarabo, C.M., Martin-Hernandez, A.M., Saiz, J.C. and Escarmis, C. (1998b) Quasispecies structure and persistence of RNA viruses. *Emerging Infectious Diseases* **4**, 521-527.
- Domingo, E., Escarmis, C., Sevilla, N., Moya, A., Elena, S.F., Quer, J., Novella, I.S. and Holland, J.J. (1996) Basic concepts in RNA virus evolution. *FASEB* **10**, 859-864.
- Domingo, E., Holland, J.J., Biebricher, C. and Eigen, M. (1995) Quasispecies: the concept and the word. In: Gibbs, A., Calisher, C. and Garcia-Arenal, F., (Eds.) *Molecular Basis of Viral Evolution*, pp. 181-191. Cambridge: Cambridge University Press.
- Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brena, P., Villanueva, N., Najera, R., et al (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance-a review. *Gene* **40**, 1-8.
- Dumoulin, F.L., Spengler, U. and Sauerbruch, T. (1997) Hepatitis C virus-associated mixed cryoglobulinaemia. *European Journal of Gastroenterology & Hepatology* **9**, 1063-1065.
- Durante, A.J. and Heptonstall, J. (1995) How many people in England and Wales risk infection from injecting drug use?. *Communicable Disease Report CDR Review*. **5**, R40-R44
- Dusheiko, G., Schmilovitz-Weiss, H., Brown, D., McOmish, F., Yap, P.L., Sherlock, McIntyre, N. and Simmonds, P. (1994a) Hepatitis C virus genotypes: an investigation of type-specific differences in geographic origin and disease. *Hepatology* **19**, 13-18.
- Dusheiko, G. and Simmonds, P. (1994b) Sequence variability of hepatitis C virus and its clinical relevance. *Journal of Viral Hepatitis* **1**, 3-15.

- Dusheiko, G.M., Khakoo, S., Soni, P. and Grellier, L. (1996) A rational approach to the management of hepatitis C infection. *BMJ* **312**, 357-364.
- Duverlie, G., Khorsi, H., Castelain, S., Jaillon, O., Izopet, J., Lunel, F., Eb, Penin, F. and Wychowski, C. (1998) Sequence analysis of the NS5A protein of European hepatitis C virus 1b isolates and relation to interferon sensitivity. *Journal of General Virology* **79**, 1373-1381.
- Eckert, K.A. and Kunkel, T.A. (1991) DNA polymerase fidelity and the polymerase chain reaction. *Genome Research* **1**, 17-24.
- Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523.
- Eigen, M. (1992) *Steps towards life*, Oxford University Press.
- Enomoto, N., Kurosaki, M., Tanaka, Y., Marumo, F. and Sato, C. (1994) Fluctuation of hepatitis C virus quasispecies in persistent infection and interferon treatment revealed by single-strand conformation polymorphism analysis. *Journal of General Virology* **75**, 1361-1369.
- Enomoto, N., Sakuma, I., Asahina, Y., Kurosaki, M., Murakami, T., Yamamoto, C., Izumi, N., Marumo, F. and Sato, C. (1995) Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b. Sensitivity to interferon is conferred by amino acid substitutions in the NS5A region. *Journal of Clinical Investigation* **96**, 224-230.
- Everitt B.S. (1997) *The Analysis of Contingency Tables*, Chapman and Hall.
- Farci, P., Alter, H.J., Govindarajan, S., Wong, D.C., Engle, R., Lesniewski, R.R., Mushahwar, I.K., Desai, S.M., Miller, R.H. and Ogata, N. (1992) Lack of protective immunity against reinfection with hepatitis C virus. *Science* **258**, 135-140.
- Farci, P., Bukh, J. and Purcell, R.H. (1997) The quasispecies of hepatitis C virus and the host immune response. *Springer Seminars in Immunopathology* **19**, 5-26.

- Farci, P., Shimoda, A., Wong, D., Cabezon, T., De Gioannis, D., Strazzera, A., Shimizu, Y., Shapiro, M., Alter, H.J. and Purcell, R.H. (1996) Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. *PNAS* **93**, 15394-15399.
- Felsenstein, J. (1993) PHYLIP (phylogeny inference package). 3.5c. Department of Genetics, University of Washington, Seattle, USA.
- Fipaldini, C., Bellei, B. and La Monica, N. (1999) Expression of hepatitis C virus cDNA in human hepatoma cell line mediated by a hybrid baculovirus-HCV vector. *Virology* **255**, 302-311.
- Fodde, R. and Losekoot, M. (1994) Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Human Mutation* **3**, 83-94.
- Fournier, C., Sureau, C., Coste, J., Ducos, J., Pageaux, G., Larrey, D., Domergue, J. and Maurel, P. (1998) In vitro infection of adult normal human hepatocytes in primary culture by hepatitis C. *Journal of General Virology* **79**, 2367-2374.
- Fujimura, Y., Ishimoto, S., Shimoyama, T., Narita, N., Kuze, Y., Yoshioka, A., Fukui, H., Tanaka, T., Tsuda, F., Okamoto, H., Miyakawa, Y. and Mayumi, M. (1996) Genotypes and multiple infections with hepatitis C virus in patients with haemophilia A in Japan. *Journal of Viral Hepatitis* **3**, 79-84.
- Fukuda, Y., Nakano, I., Katano, Y., Toyoda, H., Imoto, M., Takamatsu, J., Saito and Hayakawa, T. (1998) Assessment and treatment of liver disease in Japanese haemophilia patients. *Haemophilia* **4**, 595-600.
- Fukuma, T., Enomoto, N., Marumo, F. and Sato, C. (1998) Mutations in the interferon-sensitivity determining region of hepatitis C virus and transcriptional activity of the nonstructural region 5A protein. *Hepatology* **28**, 1147-1153.
- Garcia, F., Roldan, C., Garcia, F.J., Hernandez, J., Garcia-Valdecasas, J., Bernal, M.C., Piedrola, G. and Maroto, M.C. (1998) Subtype distribution among intravenous drug users with chronic type C hepatitis in southern Spain. *Microbios* **95**, 15-24.

- Gomez, J., Martell, M., Quer, J., Cabot, B. and Esteban, J.I. (1999) Hepatitis C viral quasispecies. *Journal of Viral Hepatitis* **6**, 3-16.
- Gonzalez-Peralta, R.P., Liu, W.Z., Davis, G.L., Qian, K.P. and Lau, J.Y. (1997) Modulation of hepatitis C virus quasispecies heterogeneity by interferon-alpha and ribavirin therapy. *Journal of Viral Hepatitis* **4**, 99-106.
- Gretch, D.R., Polyak, S.J., Wilson, J.J., Carithers, R.L.J., Perkins, J.D. and Corey (1996) Tracking hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. *Journal of Virology* **70**, 7622-7631.
- Guadagnino, V., Stroffolini, T., Rapicetta, M., Costantino, A., Kondili, L.A., Menniti-Ippolito, F., Caroleo, B., Costa, C., Griffo, G., Loiacono, L., Pisani, V., Foca, A. and Piazza, M. (1997) Prevalence, risk factors and genotype distribution of hepatitis C virus infection in the general population: a community-based survey in southern Italy. *Hepatology* **26**, 1006-1011.
- Guldborg, P., Nedergaard, T., Nielsen, H.J., Olsen, A.C., Ahrenkiel, V. and Zeuthen (1997) Single-step DGGE-based mutation scanning of the p53 gene: application to genetic diagnosis of colorectal cancer. *Human Mutation* **9**, 348-355.
- Gwack, Y., Kim, D.W., Han, J.H. and Choe, J. (1997) DNA helicase activity of the hepatitis C virus nonstructural protein 3. *European Journal of Biochemistry* **250**, 47-54.
- Hayashi, J., Kishihara, Y., Yamaji, K., Furusyo, N., Yamamoto, T., Pae, Y., Etoh, Ikematsu, H. and Kashiwagi, S. (1997) Hepatitis C viral quasispecies and liver damage in patients with chronic hepatitis C virus infection. *Hepatology* **25**, 697-701.
- Hayashi, N. and Mita, E. (1999) Correlation between the efficacy of interferon and HCV gene. *Asian Medical Journal* **42** (1): 1-7.
- Hellen, C.U.T. and Pestova, T.V. (1999) Translation of hepatitis C virus RNA. *Journal of Viral Hepatitis* **6**, 79-87.

- Higashi, Y., Kakumu, S., Yoshioka, K., Wakita, T., Mizokami, M., Ohba, K., Ito, Ishikawa, T., Takayanagi, M. and Nagai, Y. (1993) Dynamics of genome change in the E2/NS1 region of hepatitis C virus in vivo. *Virology* **197**, 659-668.
- Holland, J. (1991) Defective viral genomes. In: Fields, B.N. and Knipe, D.N., (Eds.) *Fundamental Virology*, pp. 151-165. New York: Raven Press.
- Honda, M., Kaneko, S., Sakai, A., Unoura, M., Murakami, S. and Kobayashi, K. (1994) Degree of diversity of hepatitis C virus quasispecies and progression of liver disease. *Hepatology* **20**, 1144-1151.
- Idilman, R., De Maria, N., Colantoni, A. and Van Thiel, D.H. (1998) Pathogenesis of hepatitis B and C-induced hepatocellular carcinoma. *Journal of Viral Hepatitis* **5**, 285-299.
- Isaacson, A.H., Davis, G.L. and Lau, J.Y. (1997) Should we test hepatitis C virus genotype and viraemia level in patients with chronic hepatitis C?. *Journal of Viral Hepatitis* **4**, 285-292.
- Isobe, K., Imoto, M., Fukuda, Y., Koyama, Y., Nakano, I., Hayakawa, T. and Takamatsu, J. (1995) Hepatitis C virus infection and genotypes in Japanese hemophiliacs. *Liver* **15**, 131-134.
- Ito, T. and Lai, M.M. (1997) Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *Journal of Virology* **71**, 8698-8706.
- Ito, T., Tahara, S.M. and Lai, M.M. (1998) The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site. *Journal of Virology* **72**, 8789-8796.
- Jarvis, L.M., Watson, H.G., McOmish, F., Peutherer, J.F., Ludlam, C.A. and Simmonds, P. (1994) Frequent reinfection and reactivation of hepatitis C virus genotypes in multitransfused hemophiliacs. *Journal of Infectious Diseases* **170**, 1018-1022.

- Jeannel, D., Fretz, C., Traore, Y., Kohdjo, N., Bigot, A., Pe, G.E., Jourdan, Kourouma, K., Maertens, G., Fumoux, F., Fournel, J.J. and Stuyver, L. (1998) Evidence for high genetic diversity and long-term endemicity of hepatitis C virus genotypes 1 and 2 in West Africa. *Journal of Medical Virology* **55**, 92-97.
- Kaito, M., Watanabe, S., Tsukiyama-Kohara, K., Yamaguchi, K., Kobayashi, Y., Konishi, M., Yokoi, M., Ishida, S., Suzuki, S. and Kohara, M. (1994) Hepatitis C virus particle detected by immunoelectron microscopic study. *Journal of General Virology* **75**, 1755-1760.
- Kanistanon, D., Neelamek, M., Dharakul, T. and Songsivilai, S. (1997) Genotypic distribution of hepatitis C virus in different regions of Thailand. *Journal of Clinical Microbiology* **35**, 1772-1776.
- Kao, J.H., Chen, P.J., Lai, M.Y. and Chen, D.S. (1993) Superinfection of heterologous hepatitis C virus in a patient with chronic type C hepatitis. *Gastroenterology* **105**, 583-587.
- Kao, J.H., Chen, P.J., Wang, J.T., Yang, P.M., Lai, M.Y., Wang, T.H. and Chen, D.S. (1996) Superinfection by homotypic virus in hepatitis C virus carriers: studies on patients with post-transfusion hepatitis. *Journal of Medical Virology* **50**, 303-308.
- Kato, N., Ikeda, M., Sugiyama, K., Mizutani, T., Tanaka, T. and Shimotohno, K. (1998) Hepatitis C virus population dynamics in human lymphocytes and hepatocytes infected in vitro. *Journal of General Virology* **79**, 1859-1869.
- Kato, N., Lan, K.H., Ono-Nita, S.K., Shiratori, Y. and Omata, M. (1997) Hepatitis C virus nonstructural region 5A protein is a potent transcriptional activator. *Journal of Virology* **71**, 8856-8859.
- Kato, N., Ootsuyama, Y., Sekiya, H., Ohkoshi, S., Nakazawa, T., Hijikata, M. and Shimotohno, K. (1994) Genetic drift in hypervariable region 1 of the viral genome in persistent hepatitis C virus infection. *Journal of Virology* **68**, 4776-4784.
- Kato, N., Sekiya, H., Ootsuyama, Y., Nakazawa, T., Hijikata, M., Ohkoshi, S. and Shimotohno, K. (1993) Humoral immune response to hypervariable region 1 of the

- putative envelope glycoprotein (gp70) of hepatitis C virus. *Journal of Virology* **67**, 3923-3930.
- Kiyosawa, K. (1997) The value of hepatitis C virus genotyping to epidemiological and clinical studies. *Journal of Gastroenterology & Hepatology* **12**, 623-624.
- Koizumi, K., Enomoto, N., Kurosaki, M., Murakami, T., Izumi, N., Marumo, F. and Sato, C. (1995) Diversity of quasispecies in various disease stages of chronic hepatitis C virus infection and its significance in interferon treatment. *Hepatology* **22**, 30-35.
- Korenaga, M., Hino, K., Okazaki, M., Okuda, M. and Okita, K. (1997) Differences in hypervariable region 1 quasispecies between immune complexed and non-immune complexed hepatitis C virus particles. *Biochemical & Biophysical Research Communications* **240**, 677-682.
- Koziel, M.J. (1997) The role of immune responses in the pathogenesis of hepatitis C virus infection. *Journal of Viral Hepatitis* **4 Suppl 2**, 31-41.
- Kumar, S., Tamura, K. and Nei, M. (1993) MEGA: molecular evolutionary genetics. Philadelphia: The Pennsylvania State University.
- Kuo, G., Choo, Q.L., Alter, H.J., Gitnick, G.L., Redeker, A.G., Purcell, R.H., Miyamura, T., Dienstag, J.L., Alter, M.J. and Stevens, C.E. (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244**, 362-364.
- Kurosaki, M., Enomoto, N., Murakami, T., Sakuma, I., Asahina, Y., Yamamoto, C., Ikeda, T., Tozuka, S., Izumi, N., Marumo, F. and Sato, C. (1997) Analysis of genotypes and amino acid residues 2209 to 2248 of the NS5A region of hepatitis C virus in relation to the response to interferon-beta therapy. *Hepatology* **25**, 750-753.
- Lawal, Z., Petrik, J., Wong, V.S., Alexander, G.J. and Allain, J.P. (1997) Hepatitis C virus genomic variability in untreated and immunosuppressed patients. *Virology* **228**, 107-111.

- Le Guen, B., Squadrito, G., Nalpas, B., Berthelot, P., Pol, S. and Brechot, C. (1997) Hepatitis C virus genome complexity correlates with response to interferon therapy: a study in French patients with chronic hepatitis C. *Hepatology* **25**, 1250-1254.
- Lechner, S., Rispeter, K., Meisel, H., Kraas, W., Jung, G., Roggendorf, M. and Zibert, A. (1998) Antibodies directed to envelope proteins of hepatitis C virus outside of hypervariable region 1. *Virology* **243**, 313-321.
- Lee, J.H., Stripf, T., Roth, W.K. and Zeuzem, S. (1997) Non-isotopic detection of hepatitis C virus quasispecies by single strand conformation polymorphism. *Journal of Medical Virology* **53**, 245-251.
- Leone, F., Zylberberg, H., Squadrito, G., Le Guen, B., Berthelot, P., Pol, S. and Brechot, C. (1998) Hepatitis C virus (HCV) hypervariable region 1 complexity does not correlate with severity of liver disease, HCV type, viral load or duration of infection. *Journal of Hepatology* **29**, 689-694.
- Lerman, L.S. and Silverstein, K. (1987) Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods in Enzymology* **155**, 482-501.
- Lin, H.J., Shi, N., Mizokami, M. and Hollinger, F.B. (1992) Polymerase chain reaction assay for hepatitis C virus RNA using a single tube for reverse transcription and serial rounds of amplification with nested primer pairs. *Journal of Medical Virology* **38**, 220-225.
- Lin, R., Liddle, C., Byth, K. and Farrell, G.C. (1996) Virus and host factors are both important determinants of response to interferon treatment among patients with chronic hepatitis C. *Journal of Viral Hepatitis* **3**, 85-96.
- Littlejohn, M., Locarnini, S. and Bartholomeusz, A. (1998) Targets for inhibition of hepatitis C virus replication. *Antiviral Therapy* **3**, 83-91.
- Lohmann, V., Korner, F., Herian, U. and Bartenschlager, R. (1997) Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of

- amino acid sequence motifs essential for enzymatic activity. *Journal of Virology* **71**, 8416-8428.
- Lohmann, V., Roos, A., Korner, F., Koch, J.O. and Bartenschlager, R. (1998) Biochemical and kinetic analyses of NS5B RNA-dependent RNA polymerase of the hepatitis C virus. *Virology* **249**, 108-118.
- Lopez-Labrador, F.X., Ampurdanes, S., Forns, X., Castells, A., Saiz, J.C., Costa, Bruix, J., Sanchez, T.J., Jimenez, d.A.M. and Rodes, J. (1997) Hepatitis C virus (HCV) genotypes in Spanish patients with HCV infection: relationship between HCV genotype 1b, cirrhosis and hepatocellular carcinoma. *Journal of Hepatology* **27**, 959-965.
- Lopez-Labrador, F.X., Ampurdanes, S., Gimenez-Barcons, M., Guilera, M., Costa, J., Jimenez, d.A.M., Sanchez-Tapias, J.M., Rodes, J. and Saiz, J.C. (1999) Relationship of the genome complexity of hepatitis C virus with liver disease severity and response to interferon in patients with chronic HCV genotype 1b infection. *Hepatology* **29**, 897-903.
- Lu, M., Funsch, B., Wiese, M. and Roggendorf, M. (1995) Analysis of hepatitis C virus quasispecies populations by temperature gradient gel electrophoresis. *Journal of General Virology* **76**, 881-887.
- Lunel, F. and Musset, L. (1998) Hepatitis C virus infection and cryoglobulinemia. *Journal of Hepatology* **29**, 848-855.
- Maggi, F., Fornai, C., Morrica, A., Vatteroni, M.L., Giorgi, M., Marchi, S., Ciccorossi, P., Bendinelli, M. and Pistello, M. (1999) Divergent evolution of hepatitis C virus in liver and peripheral blood mononuclear cells of infected patients. *Journal of Medical Virology* **57**, 57-63.
- Maggi, F., Fornai, C., Vatteroni, M.L., Giorgi, M., Morrica, A., Pistello, M., Cammarota, G., Marchi, S., Ciccorossi, P., Bionda, A. and Bendinelli, M. (1997) Differences in hepatitis C virus quasispecies composition between liver, peripheral blood mononuclear cells and plasma. *Journal of General Virology* **78**, 1521-1525.

- Major, M.E. and Feinstone, S.M. (1997) The molecular virology of hepatitis C. [Review]. *Hepatology* **25**, 1527-1538.
- Malet, I., Wychowski, C., Huraux, J.M., Agut, H. and Cahour, A. (1998) Yellow fever 5' noncoding region as a potential element to improve hepatitis C virus production through modification of translational control. *Biochemical & Biophysical Research Communications* **253**, 257-264.
- Manzin, A., Solforosi, L., Petrelli, E., Macarri, G., Tosone, G., Piazza, M. and Clementi, M. (1998) Evolution of hypervariable region 1 of hepatitis C virus in primary infection. *Journal of Virology* **72**, 6271-6276.
- Martell, M., Esteban, J.I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J. and Gomez, J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *Journal of Virology* **66**, 3225-3229.
- Marusawa, H., Hijikata, M., Chiba, T. and Shimotohno, K. (1999) Hepatitis C virus core protein inhibits Fas- and tumor necrosis factor alpha-mediated apoptosis via NF-kB activation. *Journal of Virology* **73**, 4713-4720.
- McAllister, J., Casino, C., Davidson, F., Power, J., Lawlor, E., Yap, P.L., Simmonds, P. and Smith, D.B. (1998) Long-term evolution of the hypervariable region of hepatitis C virus in a common-source-infected cohort. *Journal of Virology* **72**, 4893-4905.
- Mellor, J., Holmes, E.C., Jarvis, L.M., Yap, P.L. and Simmonds, P. (1995) Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. The International HCV Collaborative Study Group. *Journal of General Virology* **76**, 2493-2507.
- Melnick, J.L. (1996) Fields, B.N., Knipe, D.M. and Howley, P.M., (Eds.) *Fields Virology*, 3rd edn. pp. 655-712. Philadelphia: Lippincott-Raven Publishers.

- Mihm, S., Fayyazi, A., Hartmann, H. and Ramadori, G. (1997) Analysis of histopathological manifestations of chronic hepatitis C virus infection with respect to virus genotype. *Hepatology* **25**, 735-739.
- Moribe, T., Hayashi, N., Kanazawa, Y., Mita, E., Fusamoto, H., Negi, M., Kaneshige, T., Igimi, H., Kamada, T. and Uchida, K. (1995) Hepatitis C viral complexity detected by single-strand conformation polymorphism and response to interferon therapy. *Gastroenterology* **108**, 789-795.
- Myers, R.M., Maniatis, T. and Lerman, L.S. (1987) Detection and localization of single base changes by denaturing gradient gel electrophoresis. *Methods in Enzymology* **155**, 501-527.
- Nagasaka, A., Hige, S., Tsunematsu, I., Yoshida, J., Sasaki, Y., Matsushima, T. and Asaka, M. (1996) Changes in hepatitis C virus quasispecies and density populations in patients before and after interferon therapy. *Journal of Medical Virology* **50**, 214-220.
- Navas, S., Martin, J., Quiroga, J.A., Castillo, I. and Carreno, V. (1998) Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients. *Journal of Virology* **72**, 1640-1646.
- Neville, J.A., Prescott, L.E., Bhattacharjee, V., Adams, N., Pike, I., Rodgers, B., El-Zayadi, A., Hamid, S., Dusheiko, G.M., Saeed, A.A., Haydon, G.H. and Simmonds, P. (1997) Antigenic variation of core, NS3, and NS5 proteins among genotypes of hepatitis C virus. *Journal of Clinical Microbiology* **35**, 3062-3070.
- Ni, Y.H., Chang, M.H., Chen, P.J., Hsu, H.Y., Lu, T.W. and Lin, K.H. (1999) Decreased diversity of Hepatitis C virus quasispecies during bone marrow transplantation. *Journal of Medical Virology* **58**, 132-138.
- Ni, Y.H., Chang, M.H., Chen, P.J., Lin, H.H. and Hsu, H.Y. (1997) Evolution of hepatitis C virus quasispecies in mothers and infants infected through mother-to-infant transmission. *Journal of Hepatology* **26**, 967-974.

- Nowak, M.A., Anderson, R.M., McLean, A.R., Wolfs, T.F., Goudsmit, J. and May, R.M. (1991) Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963-969.
- Odeberg, J., Yun, Z., Sonnerborg, A., Bjoro, K., Uhlen, M. and Lundeberg, J. (1997) Variation of hepatitis C virus hypervariable region 1 in immunocompromised patients. *Journal of Infectious Diseases* **175**, 938-943.
- Ohno, O., Mizokami, M., Wu, R.R., Saleh, M.G., Ohba, K., Orito, E., Mukaide, M., Williams, R. and Lau, J.Y. (1997) New hepatitis C virus (HCV) genotyping system that allows for identification of HCV genotypes 1a, 1b, 2a, 2b, 3a, 3b, 4, 5a, and 6a. *Journal of Clinical Microbiology* **35**, 201-207.
- Okamoto, H., Mishiro, S., Tokita, H., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1994) Superinfection of chimpanzees carrying hepatitis C virus of genotype II/1b with that of genotype III/2a or I/1a. *Hepatology* **20**, 1131-1136.
- Okamoto, H., Sugiyama, Y., Okada, S., Kurai, K., Akahane, Y., Sugai, Y., Tanaka, Sato, K., Tsuda, F. and Miyakawa, Y. (1992) Typing hepatitis C virus by polymerase chain reaction with type-specific primers: application to clinical surveys and tracing infectious sources. *Journal of General Virology* **73**, 673-679.
- Pawlotsky, J.M. (1998a) Hepatitis C virus infection: virus/host interactions. [Review]. *Journal of Viral Hepatitis* **5 Suppl 1**, 3-8.
- Pawlotsky, J.M., Germanidis, G., Neumann, A.U., Pellerin, M., Frainais, P.O. and Dhumeaux, D. (1998b) Interferon resistance of hepatitis C virus genotype 1b: relationship to nonstructural 5A gene quasispecies mutations. *Journal of Virology* **72**, 2795-2805.
- Pawlotsky, J.M., Pellerin, M., Bouvier, M., Roudot-Thoraval, F., Germanidis, G., Bastie, A., Darthuy, F., Remire, J., Soussy, C.J. and Dhumeaux, D. (1998c) Genetic complexity of the hypervariable region 1 (HVR1) of hepatitis C virus (HCV): influence on the characteristics of the infection and responses to interferon alfa

- therapy in patients with chronic hepatitis C. *Journal of Medical Virology* **54**, 256-264.
- Pawlotsky, J.M., Tsakiris, L., Roudot-Thoraval, F., Pellet, C., Stuyver, L., Duval, J. and Dhumeaux, D. (1995) Relationship between hepatitis C virus genotypes and sources of infection in patients with chronic hepatitis C. *Journal of Infectious Diseases* **171**, 1607-1610.
- Perrault, J. (1981) Origin and replication of defective interfering particles. [Review]. *Current Topics in Microbiology & Immunology* **93**, 151-207.
- Picchio, G.R., Nakatsuno, M., Boggiano, C., Sabbe, R., Corti, M., Daruich, J., Perez-Bianco, R., Tezanos-Pinto, M., Kokka, R., Wilber, J. and Mosier, D. (1997) Hepatitis C (HCV) genotype and viral titer distribution among Argentinean hemophilic patients in the presence or absence of human immunodeficiency virus (HIV) co-infection. *Journal of Medical Virology* **52**, 219-225.
- Pipan, C., Amici, S., Astori, G., Ceci, G.P. and Botta, G.A. (1996) Vertical transmission of hepatitis C virus in low-risk pregnant women. *European Journal of Clinical Microbiology & Infectious Diseases* **15**, 116-120.
- Pohjanpelto, P., Lappalainen, M., Widell, A., Asikainen, K. and Paunio, M. (1996) Hepatitis C genotypes in Finland determined by RFLP. *Clinical & Diagnostic Virology* **7**, 7-16.
- Polyak, S.J., Faulkner, G., Carithers, R.L.J., Corey, L. and Gretch, D.R. (1997) Assessment of hepatitis C virus quasispecies heterogeneity by gel shift analysis: correlation with response to interferon therapy. *Journal of Infectious Diseases* **175**, 1101-1107.
- Polyak, S.J., McArdle, S., Liu, S.L., Sullivan, D.G., Chung, M., Hofgartner, W.T., Carithers, R.L.J., McMahon, B.J., Mullins, J.I., Corey, L. and Gretch, D.R. (1998) Evolution of hepatitis C virus quasispecies in hypervariable region 1 and the putative interferon sensitivity-determining region during interferon therapy and natural infection. *Journal of Virology* **72**, 4288-4296.

- Polyak, S.J., Paschal, D.M., McArdle, S., Gale, M.J.J., Moradpour, D. and Gretch, D.R. (1999) Characterization of the effects of hepatitis C virus nonstructural 5A protein expression in human cell lines and on interferon-sensitive virus replication. *Hepatology* **29**, 1262-1271.
- Pujol, F.H., Devesa, M., Loureiro, C.L., Capriles, F. and Liprandi, F. (1998) Turnover of hepatitis C virus genotypes in hemodialysis patients. *Archives of Virology* **143**, 823-827.
- Ramsay, M.E., Balogun, M.A., Collins, M. and Balraj, V. (1998) Laboratory surveillance of hepatitis C virus infection in England and Wales: 1992 to 1996. *Communicable Disease and Public Health* **1**, 89-94.
- Rapicetta, M., Argentini, C., Dettori, S., Spada, E., Pellizzer, G. and Gandin, C. (1998) Molecular heterogeneity and new subtypes of HCV genotype 4. *Research in Virology* **149**, 293-297.
- Ray, R.B., Meyer, K. and Ray, R. (1996) Suppression of apoptotic cell death by hepatitis C virus core protein. *Virology* **226**, 176-182.
- Ray, R.B., Meyer, K., Steele, R., Shrivastava, A., Aggarwal, B.B. and Ray, R. (1998a) Inhibition of tumor necrosis factor (TNF- α)- mediated apoptosis by hepatitis C virus core protein. *Journal of Biological Chemistry* **273**, 2256-2259.
- Ray, R.B., Steele, R., Meyer, K. and Ray, R. (1998b) Hepatitis C virus core protein represses p21WAF1/Cip1/Sid1 promoter activity. *Gene* **208**, 331-336.
- Ray, S.C., Wang, Y.M., Laeyendecker, O., Ticehurst, J.R., Villano, S.A. and Thomas, D.L. (1999) Acute hepatitis C virus structural gene sequences as predictors of persistent viremia: hypervariable region 1 as a decoy. *Journal of Virology* **73**, 2938-2946.
- Rice, C.M. (1996) Fields, B.N., Knipe, D.M. and Howley, P.M., (Eds.) *Fields Virology*, 3rd edn. pp. 931-959. Philadelphia: Lippincott-Raven Publishers.

- Rispeter, K., Lu, M., Zibert, A., Wiese, M., de Oliveira, J.M. and Roggendorf, M. (1998) The "interferon sensitivity determining region" of hepatitis C virus is a stable sequence element. *Journal of Hepatology* **29**, 352-361.
- Ruggieri, A., Argentini, C., Kouruma, F., Chionne, P., D'Ugo, E., Spada, E., Dettori, S., Sabbatani, S. and Rapicetta, M. (1996) Heterogeneity of hepatitis C virus genotype 2 variants in West Central Africa (Guinea Conakry). *Journal of General Virology* **77**, 2073-2076.
- Saiz, J.C., Lopez-Labrador, F.X., Ampurdanes, S., Dopazo, J., Forns, X., Sanchez-Tapias, J.M. and Rodes, J. (1998) The prognostic relevance of the nonstructural 5A gene interferon sensitivity determining region is different in infections with genotype 1b and 3a isolates of hepatitis C virus. *Journal of Infectious Diseases* **177**, 839-847.
- Sangar, D.V. and Carroll, A.R. (1998) A tale of two strands: reverse-transcriptase polymerase chain reaction detection of hepatitis C virus replication. *Hepatology* **28**, 1173-1176.
- Sheffield, V.C., Cox, D.R., Lerman, L.S. and Myers, R.M. (1989) Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *PNAS* **86**, 232-236.
- Shimizu, Y.K., Hijikata, M., Iwamoto, A., Alter, H.J., Purcell, R.H. and Yoshikura, H. (1994) Neutralizing antibodies against hepatitis C virus and the emergence of neutralization escape mutant viruses. *Journal of Virology* **68**, 1494-1500.
- Silini, E., Bono, F., Cividini, A., Cerino, A., Maccabruni, A., Tinelli, C., Bruno, S., Bellobuono, A. and Mondelli, M. (1995) Molecular epidemiology of hepatitis C virus infection among intravenous drug users. *Journal of Hepatology* **22**, 691-695.
- Simmonds, P. (1995) Variability of hepatitis C virus. *Hepatology* **21**, 570-583.
- Simmonds, P. (1998) Variability of the hepatitis C virus genome. *Current Studies in Hematology & Blood Transfusion* 38-63.

- Simmonds, P., Holmes, E.C., Cha, T.A., Chan, S.W., McOmish, F., Irvine, B., Beall, Yap, P.L., Kolberg, J. and Urdea, M.S. (1993) Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *Journal of General Virology* **74**, 2391-2399.
- Simmonds, P., Mellor, J., Sakuldamrongpanich, T., Nuchaprayoon, C., Tanprasert, Holmes, E.C. and Smith, D.B. (1996) Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *Journal of General Virology* **77**, 3013-3024.
- Smith, D.B., McAllister, J., Casino, C. and Simmonds, P. (1997) Virus 'quasispecies': making a mountain out of a molehill?. *Journal of General Virology* **78**, 1511-1519.
- Smith, D.B., Mellor, J., Jarvis, L.M., Davidson, F., Kolberg, J., Urdea, M., Yap, PL and Simmonds, P. (1995) Variation of the hepatitis C virus 5' non-coding region: implications for secondary structure, virus detection and typing. The International HCV Collaborative Study Group. *Journal of General Virology* **76**, 1749-1761.
- Smith, D.B. and Simmonds, P. (1997) Review: molecular epidemiology of hepatitis C virus. *Journal of Gastroenterology & Hepatology* **12**, 522-527.
- Smuts, H.E. and Kannemeyer, J. (1995) Genotyping of hepatitis C virus in South Africa. *Journal of Clinical Microbiology* **33**, 1679-1681.
- Stark, K., Schreier, E., Muller, R., Wirth, D., Driesel, G. and Bienzle, U. (1995) Prevalence and determinants of anti-HCV seropositivity and of HCV genotype among intravenous drug users in Berlin. *Scandinavian Journal of Infectious Diseases* **27**, 331-337.
- Steinhauer, D.A. and Holland, J.J. (1987) Rapid evolution of RNA viruses. *Annual Review of Microbiology* **41**, 409-433.
- Stuyver, L., Wyseur, A., van Arnhem, W., Hernandez, F. and Maertens, G. (1996) Second-generation line probe assay for hepatitis C virus genotyping. *Journal of Clinical Microbiology* **34**, 2259-2266.

- Sullivan, D.G., Wilson, J.J., Carithers, R.L.J., Perkins, J.D. and Gretch, D.R. (1998) Multigene tracking of hepatitis C virus quasispecies after liver transplantation: correlation of genetic diversification in the envelope region with asymptomatic or mild disease patterns. *Journal of Virology* **72**, 10036-10043.
- Tagariello, G., Pontisso, P., Davoli, P.G., Ruvoletto, M.G., Traldi, A. and Alberti (1995) Hepatitis C virus genotypes and severity of chronic liver disease in haemophiliacs. *British Journal of Haematology* **91**, 708-713.
- Taylor, D.R., Shi, S.T., Romano, P.R., Barber, G.N. and Lai, M.M. (1999) Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* **285**, 107-109.
- Tokita, H., Okamoto, H., Iizuka, H., Kishimoto, J., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1998) The entire nucleotide sequences of three hepatitis C virus isolates in genetic groups 7-9 and comparison with those in the other eight genetic groups. *Journal of General Virology* **79**, 1847-1857.
- Tokita, H., Okamoto, H., Luengrojanakul, P., Vareesangthip, K., Chainuvati, T., Iizuka, H., Tsuda, F., Miyakawa, Y. and Mayumi, M. (1995) Hepatitis C virus variants from Thailand classifiable into five novel genotypes in the sixth (6b), seventh (7c, 7d) and ninth (9b, 9c) major genetic groups. *Journal of General Virology* **76**, 2329-2335.
- Tokita, H., Okamoto, H., Tsuda, F., Song, P., Nakata, S., Chosa, T., Iizuka, H., Mishiro, S., Miyakawa, Y. and Mayumi, M. (1994) Hepatitis C virus variants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. *PNAS* **91**, 11022-11026.
- Toyoda, H., Fukuda, Y., Koyama, Y., Nakano, I., Kinoshita, M., Hadama, T., Takamatsu, J. and Hayakawa, T. (1996) Nucleotide sequence diversity of hypervariable region 1 of hepatitis C virus in Japanese hemophiliacs with chronic hepatitis C and patients with chronic posttransfusion hepatitis C. *Blood* **88**, 1488-1493.

- Toyoda, H., Fukuda, Y., Nakano, I., Katano, Y., Takayama, T., Kumada, T., Nakano, Takamatsu, J., Saito, H. and Hayakawa, T. (1998) Quasispecies nature of hepatitis C virus (HCV) in patients with chronic hepatitis C with mixed HCV subtypes. *Journal of Medical Virology* **54**, 80-85.
- Toyoda, H., Kumada, T., Nakano, S., Takeda, I., Sugiyama, K., Osada, T., Kiriya, S., Sone, Y., Kinoshita, M. and Hadama, T. (1997) Quasispecies nature of hepatitis C virus and response to alpha interferon: significance as a predictor of direct response to interferon. *Journal of Hepatology* **26**, 6-13.
- Tsuchihara, K., Hijikata, M., Fukuda, K., Kuroki, T., Yamamoto, N. and Shimotohno (1999) Hepatitis C virus core protein regulates cell growth and signal transduction pathway transmitting growth stimuli. *Virology* **258**, 100-107.
- Tuveri, R., Rothschild, C., Pol, S., Reijasse, D., Persico, T., Gazengel, C., Brechot, C. and Thiers, V. (1997) Hepatitis C virus genotypes in French haemophiliacs: kinetics and reappraisal of mixed infections. *Journal of Medical Virology* **51**, 36-41.
- Unlinked Anonymous HIV Surveys Steering Group. (1996) Unlinked anonymous HIV seroprevalence monitoring programme in England and Wales. Data to the end of 1995. *Department of Health, Public Health Laboratory Service, Institute of Child Health, London*.
- van Doorn, L.J., Capriles, I., Maertens, G., DeLeys, R., Murray, K., Kos, T., Schellekens, H. and Quint, W. (1995) Sequence evolution of the hypervariable region in the putative envelope region E2/NS1 of hepatitis C virus is correlated with specific humoral immune responses. *Journal of Virology* **69**, 773-778.
- Vitale, F., Villafrate, M.R., Viviano, E., Perna, A.M., Bonura, F., Di, B., MA, Mazzola, G., Colletti, P., Prestileo, T. and Romano, N. (1998) Distribution of hepatitis C virus genotypes among intravenous drug users. A ten-year study in Palermo, Sicily. *New Microbiologica* **21**, 335-342.
- Wang, Y.M., Ray, S.C., Laeyendecker, O., Ticehurst, J.R. and Thomas, D.L. (1998) Assessment of hepatitis C virus sequence complexity by electrophoretic mobilities

- of both single-and double-stranded DNAs. *Journal of Clinical Microbiology* **36**, 2982-2989.
- Wansbrough-Jones, M.H., Frimpong, E., Cant, B., Harris, K., Evans, M.R. and Teo, C.G. (1998) Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana. *Transactions of the Royal Society of Tropical Medicine & Hygiene* **92**, 496-499.
- Weiner, A.J., Geysen, H.M., Christopherson, C., Hall, J.E., Mason, T.J., Saracco, G., Bonino, F., Crawford, K., Marion, C.D. and Crawford, K.A. (1992) Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections. *PNAS* **89**, 3468-3472.
- Woodward, T.M., Carlson, J., McClelland, C. and DeMartini, J.C. (1994) Analysis of lentiviral genomic variation by denaturing gradient gel electrophoresis. *Biotechniques* **17**, 366-371.
- Wright, S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97-159.
- Wright, T., Hollander, H. and Pu, X. (1994) Hepatitis C in HIV-infected patients with and without AIDS: prevalence and relationship to patient survival. *Hepatology* **20**, 1152-1155.
- Wu, C.H., Lee, M.F. and Kuo, H.S. (1997) Distribution of hepatitis C virus genotypes among blood donors in Taiwan. *Journal of Gastroenterology & Hepatology* **12**, 625-628.
- Wyatt, C.A., Andrus, L., Brotman, B., Huang, F., Lee, D.H. and Prince, A.M. (1998) Immunity in chimpanzees chronically infected with hepatitis C virus: role of minor quasispecies in reinfection. *Journal of Virology* **72**, 1725-1730.
- Zein, N.N., Rakela, J., Krawitt, E.L., Reddy, K.R., Tominaga, T. and Persing, D.H. (1996) Hepatitis C virus genotypes in the United States: epidemiology, pathogenicity, and response to interferon therapy. Collaborative Study Group. *Annals of Internal Medicine* **125**, 634-639.

- Zeuzem, S., Lee, J.H. and Roth, W.K. (1997) Mutations in the nonstructural 5A gene of European hepatitis C virus isolates and response to interferon alfa. *Hepatology* **25**, 740-744.
- Zibert, A., Kraas, W., Meisel, H., Jung, G. and Roggendorf, M. (1997) Epitope mapping of antibodies directed against hypervariable region 1 in acute self-limiting and chronic infections due to hepatitis C virus. *Journal of Virology* **71**, 4123-4127.
- Zibert, A., Schreier, E. and Roggendorf, M. (1995) Antibodies in human sera specific to hypervariable region 1 of hepatitis C virus can block viral attachment. *Virology* **208**, 653-661.

Publications and presentations relating to this thesis

Harris K. and Teo C.G. Analysis of hepatitis C virus quasispecies using denaturing gradient gel electrophoresis. Manuscript in preparation.

Harris K., Gilham C., Mortimer P.P., Teo C.G. (1999) The most prevalent hepatitis C virus genotypes in England and Wales are 3a and 1a. *Journal of Medical Virology* **58**: 127-131

Wansbrough-Jones M.H., Frimpole E., Cant B., Harris K., Evans M.R. and Teo C.G. (1998) Prevalence and genotype of hepatitis C virus infection in pregnant women and blood donors in Ghana. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **92**: 496-499.

Harris K. and Teo C.G. Hepatitis C virus genotypes in England and Wales. Oral presentation. PHLS 23rd Annual Scientific Conference, University of Warwick, September 7th-9th 1998.

Harris K., Hallett R. and Teo C.G. Extent of infection with multiple HCV genotype. Poster presentation. Vth International Conference: Current Trends in Chronically Evolving Hepatitis, Lyon, France, October 10th-11th 1997.